Estudos Avançados em Organização do Conhecimento. v.3

Organização do Conhecimento e Diversidade Cultural

Organizadores: José Augusto Chaves Guimarães e Vera Dodebei

ISKO-BRASIL

fundepecursos

Organizadores

José Augusto Chaves Guimarães
Vera Dodebei

Organização do Conhecimento e Diversidade Cultural

Marília, São Paulo, Brasil

Sociedade Brasileira de Organização do Conhecimento
(ISKO-Brasil)

Fundação para o Desenvolvimento do Ensino, Pesquisa e Extensão
(FUNDEPE)

2015

O68o Organização do conhecimento e diversidade cultural [recurso eletrônico] / José Augusto Chaves Guimarães, Vera Dodebei, organizadores. -- Marília: ISKO-Brasil ; FUNDEPE, 2015.
835 f. ; 30 cm.

ISBN: 978-85-98176-70-3

Livro digital

1. Organização do conhecimento. I. Título.

CDD 025.4

Comissão Científica

Vera Lúcia D. Louzada de Mattos Dodebei (UNIRIO)

Gercina Ângela Borém de Oliveira Lima (UFMG)

Marisa Bräscher Basílio Medeiros (UFSC)

Maria Aparecida Moura (UFMG)

Fabio Assis Pinho (UFPE)

Editoração

Isadora Victorino Evangelista

Gilberto Gomes Cândido

Rafael Aparecido Moron Semidão

Rafael Cacciolari Dalessandro

Suellen Oliveira Milani

Capa e arte gráfica

Maíra Fernandes Alencar

© A reprodução deste livro na íntegra ou em parte é permitida, desde que citados os créditos. Proibida a venda.

SUMÁRIO

APRESENTAÇÃO.....13 - 19

A DIMENSÃO EPISTEMOLÓGICA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

Epistemic, ontic, axiologic, and praxic constructs in knowledge organization research…….21 - 29
Joseph T. Tennis

Abordagens e paradigmas na organização do conhecimento.....30 - 36
Leila Cristina Weiss e Marisa Bräscher

A dimensão histórico-epistemológica da Organização do Conhecimento: contributos da taxonomia Gesneriana, Séc. XVI.....37 - 46
Andre Vieira de Freitas Araujo e Giulia Crippa

Configurações terminológicas e interdisciplinares na pesquisa em organização do conhecimento (2010-2014).....47 - 54
Leilah Santiago Bufrem, Aline Elis Arboit e Juliana Lazzarotto Freitas

A dimensão conceitual da organização do conhecimento nos congressos da NASKO: análise de conteúdo Bardiana.....55 - 72
José Augusto Chaves Guimarães, Rodrigo de Sales, André Ynada dos Santos e Daniela Fernanda de Oliveira Matos

A relação entre organização do conhecimento e ciência da informação na comunidade científica brasileira: uma investigação no âmbito da ISKO-Brasil.....73 - 84
Rodrigo de Sales

Poole, o índice e as fraturas: Indexação e publicações seriadas nos Estados Unidos do Oitocentos.....85 - 93
Gustavo Silva Saldanha e Naira Christofoletti Silveira

Os processos de representação do conhecimento arquivístico: elementos históricos e conceituais da classificação e descrição.....94 - 99
Natália Bolfarini Tognoli e Thiago Henrique Bragato Barros

Ontologias e Knowledge Organization System (KOS): aproximações e diferenças.....100 - 107
Rogério Aparecido Sá Ramalho

Estudos de genealogia acadêmica como abordagem para análise de domínio.....108 - 116
Renata Cristina Gutierres Castanha e Maria Cláudia Cabrini Grácio

A dimensão epistemológica da análise documental de conteúdo de obras de ficção na organização do conhecimento.....117 - 124
Roberta Caroline Vesú Alves e João Batista Ernesto de Moraes

As linguagens de indexação e a análise de domínio.....125 - 134
Vera Lucia Ribeiro Guim e Mariângela Spotti Lopes Fujita

Reader-Interest classification as a method of stock control: the McClellan Legacy…….....135 - 142
Daniel Martínez-Ávila e *M. P. Satija*

Terminologia, exaustividade e especificidade: uma relação conceitual.....143 - 153
Isadora Victorino Evangelista, Walter Moreira e João Batista Ernesto de Moraes

Aspectos dos metadados de assunto na catalogação de fotografias.....154 - 162
Ana Carolina Simionato

Linguagens documentárias e tratamento de recursos audiovisuais: interfaces teóricas e metodólogicas.....163 - 172
Francisnaira Cristina Ravazzi e Walter Moreira

Organização do conhecimento e design da informação: um estudo de convergência.....173 - 183
Natalia Nakano, Mariana Cantisani Padua e Maria José Vicentini Jorente)

Análisis de dominio de las dos primeras ediciones del capítulo ISKO-Brasil……..184 - 199
Exequiel Fontans e Mario Barité

Ensino em organização e representação do conhecimento: linguagens e normas.....200 - 209
Cibele Araújo Marques dos Santos, Eduardo de Abreu de Jesus e João Ricardo de Luca

A genealogia do conceito de Ciência da Informação no Brasil: uma análise discursiva a partir de periódicos fundacionais na área .....210 - 218
Larissa de Melo Lima, João Batista Ernesto de Moraes e Daniel Martínez-Ávila

A DIMENSÃO APLICADA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

A Discourse Analysis of the Semantic Web……220 - 223
D. Grant Campbell

Tecnologias da web semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados.....224 - 233
José Eduardo Santarem Segundo e Caio Saraiva Coneglian

Análise preliminar sobre a conversão de plano de classificação em vocabulário controlado.....234 - 241
Luciana Davanzo e Walter Moreira

Reengenharia de tesauro: caso do Thesagro.....242 - 250
Benildes Maculan, Gercina Lima, Ivo Pierozzi Jr. e Leandro Oliveira

Modelo conceitual de inteligência competitiva organizacional.....251 - 261
Thiciane Mary Carvalho Teixeira e Marta Lígia Pomim Valentim

Controle de vocabulário em periódicos científicos eletrônicos.....262 - 271
José Carlos Francisco dos Santos e Brígida Maria Nogueira Cervantes

As linguagens documentárias na descrição arquivística.....272 - 282
Maria de Fátima Santos de Lima e Francisco Aragão Pedroza da Cunha

As pesquisas em sistemas de organização do conhecimento nas conferências DCMI.....283 - 294
Felipe Augusto Arakaki, Plácida Leopoldina Ventura Amorim da Costa Santos e Rachel Cristina Vesú Alves

Modelos conceituais e representação descritiva da informação.....295 - 302
Elisabete Gonçalves de Souza e Wellington Freire Cunha Costa

Nanopublicações e indexação: experimento metodológico.....303 - 317
Lorena Tavares de Paula e Maria Aparecida Moura

Aplicação de ontologias no processo de recuperação da informação aplicado em ambientes acadêmicos.....318 - 328
Caio Saraiva Coneglian,Elvis Fusco e José Eduardo Santarém Segundo

Atuação bibliotecária na análise de assunto de teses nas dimensões teóricas da catalogação de assunto e indexação.....329 - 337
Roberta Cristina Dal'Evedove Tartarotti, Paula Regina Dal'Evedove e Mariângela Spotti Lopes Fujita

A contribuição da identificação documental para a organização do conhecimento em arquivos pessoais.....338 - 343
Gabrieli Aparecida da Fonseca e Sonia Maria Troitiño Rodriguez

Análise documental de conteúdo e o percurso gerativo de sentido: na representação do documento de arquivo.....344 - 354
Gilberto Gomes Cândido, João Batista Ernesto de Moraes e Deise Sabbag

Contribuições do percurso gerativo de sentido para a leitura documentária de textos narrativos de ficção.....355 - 364
Deise Sabbag e João Batista Ernesto de Moraes

O modelo teórico do percurso gerativo de sentido para a elaboração de resumos de textos científicos: uma releitura do produto informacional resumo.....365 - 372
Roberta Caroline Vesú Alves e João Batista Ernesto de Moraes

El control de vocabulario y su aporte a la investigación: el caso de las ciencias agrarias de Uruguay……...…373 - 380
Lucía Simón e Mario Barité

Bens culturais rurais: relato de experiência da construção de taxonomia no contexto das fazendas históricas paulistas.....381 - 391
Mayara Cristina Bernardino,Luciana de Souza Gracioso,Maria da Graça Melo Simões e Luzia Sigoli Fernandes Costa

Harmonization of CIDOC CRM ontology in the context of archives, libraries and museums……..…392 - 398
Laís Barbudo Carrasco, Silvana Aparecida Borsetti Gregório Vidotti e Phil Manfred Thaller

Catalogación de recursos interactivos: análisis de recomendaciones y prácticas en catálogos internacionales……...399 - 412
Víctor Amor, Daniel Martínez-Ávila e Rosa San Segundo

Estrutura de metadados para banco de imagens em patologia.....413 - 424
Elisabete Gonçalves Souza, Jóice Cleide Cardoso Ennes de Souza e Elan Cardozo Paes de Almeida

Indexação por imagens: acessibilidade via OPACs imagéticos.....425 - 448
Rita de Cássia do Vale Caribé e Marcílio de Brito

Mapas conceituais interativos como instrumento didático no campo da organização da informação e do conhecimento.....449 - 460
Marilda Lopes Ginez de Lara, Gabriela Previdello e Nair Yumiko Kobashi

Análise de domínio na organização do conhecimento: explorando as relações temáticas e de citação.....461 - 468
Bruno Henrique Alves, Ely Francina Tannuri de Oliveira e Maria Cláudia Cabrini Grácio

Um estudo do domínio em organização do conhecimento por meio da análise de citação e cocitação de autores.....469 - 476

Lidyane Silva Lima, Pollyana Ágata Gomes da Rocha Custódia, Ely Francina Tannuri de Oliveira e Leilah Santiago Bufrém

Abordagens de classificação nos trabalhos apresentados ao GT2 do ENANCIB: classificação profissional e classificação não-profissional.....477 - 486
Walter Moreira e Isabela Santana de Moraes

O tema comportamento informacional no âmbito da ISKO - International Society For Knowledge Organization: estudo referente ao período de 2004 – 2014.....487 - 497
Marli Vitor da Silva e Helen de Castro Silva Casarin

A representação de assunto no escopo da arquivologia: uma análise de artigos científicos nacionais e internacionais.....498 - 506
Graziela Martins de Medeiros, Leolibia Luana Linden, Luciane Paula Vital e Marisa Bräscher

Análise da metodologia em estudos bibliométricos: uma proposição de indicadores de contexto.....507 - 517
Maria Guiomar da Cunha Frota e Ana Cláudia Ribeiro

Organização e representação do conhecimento: contribuições aos estudos métricos.....518 - 525
Cibele Araújo Marques dos Santos

Relativizando os índices h e g: estudo aplicado ao domínio dos estudos métricos.....526 - 533
Deise Deolindo Silva e Maria Cláudia Cabrini Grácio

A abordagem da análise de domínio na organização e representação do conhecimento em arquivística.....534 - 544
Cynthia Maria Kiyonaga Suenaga e Brígida Maria Nogueira Cervantes.

O Familysearch Indexing como iniciativa de crowdsourcing no contexto da organização do conhecimento.....545 - 552
Paula Carina de Araújo e José Augusto Chaves Guimarães

Estudo de ações para disponibilização de datasets governamentais em linked open data.....553 - 564
Fernando de Assis Rodrigues e Ricardo César Gonçalves Santana

Organização do conhecimento e o patrimônio industrial em São Paulo: o projeto Eletromemória.....565 - 573
Vânia Mara Alves Lima,Marcia Cristina e Carvalho Pazin Vitoriano e Cristina Hilsdorf Barbanti

A DIMENSÃO SOCIAL, CULTURAL E POLÍTICA DA ORGANIZAÇÃO E REPRESENTAÇÃO DO CONHECIMENTO

Cultural Interoperability and Knowledge Organization Systems……..575 - 606
Widad Mustafa El Hadi

Reflexões acerca do desenvolvimento de metodologia para análise de assunto.....607 – 614
Paula Regina Dal'Evedove, Roberta Cristina Dal'Evedove Tartarotti e Mariângela Spotti Lopes Fujita

Organização, representação, recuperação e acesso à informação: (re)configuração do Formato MARC21 e do BIBFRAME pelos propósitos da diversidade cultural nos ambientes informacionais digitais?.....615 - 620
Zaira Regina Zafalon e Marcela Cristina Nespoli

Música, literatura e audiovisual: as contribuições da organização do conhecimento (OC) nas relações de interseccionalidade das obras de Dorival Caymmi e Jorge Amado.....621 - 629
Fabio Assis Pinho, Francisco Arrais Nascimento, e Andréa Carla Melo Marinho

A mediação no domínio da organização do conhecimento.....630 - 642
Mona Cleide Quirino da Silva Farias, Carlos Cândido de Almeida e Daniel Martínez-Ávilla

O valor da informação e da linguagem na sociedade de consumo.....643 - 649
Luciana de Souza Gracioso

Formação acadêmica do docente em descrição arquivística.....650 - 657
Laura Maria do Rego, José Augusto Chaves Guimarães e Natália Bolfarini Tognoli

PERSPECTIVAS DA PESQUISA EM ORGANIZAÇÃO DO CONHECIMENTO NO BRASIL

Knowledge Organization and the core of Information Science: tales of big data, computing clouds and social networks……659 - 669
Renato Rocha Souza, Mauricio Barcellos Almeida e Renata Abrantes Baracho Porto

Organização do conhecimento: pesquisa e desenvolvimento.....670 - 687
Gercina Ângela Borém de Oliveira Lima

A ISKO Brasil e os grupos de pesquisa em organização do conhecimento.....688 - 695
Evelyn Goyannes Dill Orrico

Perspectivas da pesquisa em organização do conhecimento: reflexões a partir de trabalhos acadêmicos do PPGCI IBICT-UFRJ.....696 - 708
Rosali Fernandez de Souza

Perspectivas da pesquisa sobre organização do conhecimento no Brasil.....709 - 724
Leilah Santiago Bufrem

PESQUISA EM ANÁLISE DOCUMENTÁRIA NO BRASIL: A INFLUÊNCIA DE JEAN-CLAUDE GARDIN

A busca pela eficiência na representação da informação e do conhecimento – desdobramentos posteriores no pensamento de Gardin.....726 - 735
Johanna W. Smit

Jean-Claude Gardin e a análise documentária: trajetória da constituição de uma semiologia da representação.....736 - 747
Maria de Fátima Gonçalves Moreira Tálamo e Giovana Deliberali Maimone

Linguagem documentária na ótica de J.-C. Gardin.....748 - 757
Marilda L. G de Lara

Da análise documentária à terminologia: percurso teórico e metodológico.....758 - 764
Vania M. A. Lima

Ordenação de documentos: fundamentos e relações com a classificação bibliográfica.....765 - 797
Cristina D. Ortega

A ISKO-Brasil e o fortalecimento de intercâmbios e visibilidade científica em organização do conhecimento.....798 – 802
Vera Dodebei

Resumos dos trabalhos ganhadores do prêmio ISKO Brasil 2015, a serem publicados, na íntegra, na revista Knowledge Organization

A organização do conhecimento no contexto da pós-modernidade sob a perspectiva da teoria da classificação.....803
Daniel Abraão Pando e Carlos Cândido de Almeida

O princípio da garantia semântica e os estudos da Linguagem:reflexões conceituais.....804
Roger de Miranda Guedes e Maria Aparecida Moura

Interoperabilidade e Mapeamentos entre Sistemas de Organização do Conhecimento: Metatesauro-Unified Medical Language System- National Library of Medicine.....805
Julietti de Andrade e Marilda Lopes Ginez de Lara

Aspectos metodológicos nas teorias críticas em Organização do Conhecimento.....806
Daniel Martinez-Ávila, Rafael Aparecido Moron Semidão e Marcio Ferreira da Silva

Resumos dos trabalhos ganhadores do prêmio ISKO Brasil 2015, a serem publicados, na íntegra, na revista Brazilian Journal of Information Science.

Propuesta de representación del Tesauro de la UNESCO para la web semántica mediante ISO-25964……..807
Juan-Antonio Pastor-Sánchez

Modelagem Metaconceitual: aplicação à Infraestrutura de Qualidade com base na ISO 25964.....808
Catarina Felix dos S. Soares e Vera Dodebei

Tres enfoques de *bias* en organización del conocimiento: *bias* neutro, *bias* negativo y *bias* positivo……..809
Stephanie Colombo e Mario Barité

APRESENTAÇÃO

Organização do conhecimento: passado, presente e futuro em um contexto de diversidade cultural.

José Augusto Chaves Guimarães
Universidade Estadual Paulista
guima@marilia.unesp.br

A Organização do Conhecimento - OC, enquanto campo de saber interdisciplinar, espaço de convergência e diálogo teórico e aplicado (Lógica, Linguística, Documentação, Comunicação, Estatística, Psicologia, Computação, Biblioteconomia etc) e com forte relação com a Ciência da Informação, vem revelando, ao longo do tempo, uma dinâmica de consolidação científica que se explica por sua própria historicidade.

Desse modo, a abordagem da OC pressupõe uma concepção dinâmica de conhecimento que, uma vez produzido e divulgado, necessita ser organizado para que possa estar acessível à sociedade para que esta, uma vez apropriando-se desse conhecimento, possa gerar novo conhecimento que, por sua vez, será socializado e organizado, em um processo helicoidal ininterrupto.

Assim, no passado, vale dizer, até meados do século XX, duas concepções se destacam no que tange à configuração da OC: a arte, enquanto produto de um ofício decorrente de um talento especial, e a técnica, fruto do estabelecimento de regras e diretrizes que norteiam um fazer.

Nesse contexto, experiências como a organização de envelopes de argila em biblioteca da Mesopotâmia, a organização de Calímaco na Biblioteca de Alexandria, a Bibliotheca Universalis de Konrad Gessner, a taxonomia de organização dos seres de Lineu, os esquemas de classificação de Bacon, Harris, Dewey, Otlet e Ranganathan, os índices cumulativos de Poole, a indexação sistemática de Kaiser, os sistemas de indexação PRECIS e POPSI e o índice de citações de Garfield são testemunhos de um importante desenvolvimento desse campo, fornecendo bases para sua consolidação científica.

Mas é notadamente no final do século XX que se pode dizer que a OC passou a assumir uma dimensão mais acadêmica, voltada para uma construção mais científica. Tal fato, por sua vez, teve, na criação da International Society for Knowledge

Organization – ISKO, um elemento-marco, na medida em que se configurou, em âmbito internacional, um espaço para abrigar os avanços investigativos, a interlocução científica e a divulgação de um conhecimento produzido por meio de uma efetiva sistemática de publicações.

Criada em julho de 1989, na Alemanha, por Ingetraut Dahlberg, a partir de uma cisão ocorrida no âmbito da antiga Society for Classification, a ISKO conseguiu constituir, ao longo de um quarto de século, uma complexa estrutura composta por capítulos nacionais ou regionais (Alemanha/Áustria/Suíça, Brasil, Canadá/Estados Unidos, China, Espanha/Portugal, França, India, Irã, Itália, Maghreb, Polônia e Reino Unido, que congregam, hoje, mais de 600 pesquisadores, com o intuito de "promover a pesquisa, o desenvolvimento e a aplicação de sistemas de OC que possam gerar avanços nas abordagens filosófica, psicológica e semântica de ordenar o conhecimento; prover meios de comunicação entre seus membros; e funcionar como um ponto de conexão entre instituições nacionais e internacionais que se dediquem a questões relativas à organização do conhecimento" (ISKO).

Nesse contexto, a ISKO dispõe de um conjunto de publicações, de abrangência internacional, no intuito de divulgar o conhecimento científico ali gerado. Para tanto tem-se a revista Knowledge Organization, de periodicidade bimestral, a série Advances in Knowledge Organization, contendo os *proceedings* dos congressos internacionais da ISKO, realizados em anos pares, em diferentes locais do mundo e, ainda, os *proceedings* dos congressos dos capítulos da ISKO, realizados em anos ímpares.

Em termos de construção teórica, alguns aspectos merecem especial destaque no contexto da OC na atualidade, como a questão do conhecimento, seu registro e sua socialização (BUCKLAND, 1991, 1997; FROHMANN, 2004, 2009), a própria configuração epistemológica da OC enquanto campo (DAHLBERG, 1993, 1995; HJORLAND, 1993, 2003, 2008; TENNIS, 2008; LÓPEZ HUERTAS, 2008; GNOLI, 2008, 2011; DOUSA, 2009; SMIRAGLIA, 2009, 2012; ZINS, 2004; THELLEFSEN & THELLEFSEN, 2004), os aspectos culturais que vêm impactando o campo (HUDON, 1997, 1999; OLSON, 2002; BEGHTOL, 2002, 2005; GARCÍA GUTIÉRREZ, 2002; HJORLAND, 2008; MILANI & GUIMARÃES, 2011; CAMPBELL, 2009), e a questão aplicada, envolvendo aspectos relativos às linguagens de representação e o impacto tecnológico (ANDERSEN, 2002; BEAK & OLSON, 2011; BUCHEL & HILL, 2009;

EDWARDS, 2011; GREEN & FALLGREN, 2007; GREEN, 2011; HOFFMAN, 2009; KIPP, 2007).

A vista de tal cenário, a ISKO-Brasil, desde seu primeiro congresso, realizado em 2011,em Brasília, reiterando no segundo congresso, realizado no Rio de Janeiro em 2013 e, nesta oportunidade, no âmbito de seu terceiro congresso, em Marília, em 2015, vem abordando suas discussões a partir de uma tríade temática composta pelas dimensões epistemológica, voltada para e a consolidação de teorias, metodologias, paradigmas, escolas de pensamento; aplicada, fortemente ligada ao avanço tecnológico e ao de desenvolvimento de novos instrumentos; e cultural, voltada para os contextos, as comunidades e os sujeitos que interagem na OC.

Nessa ótica tridimensional, alguns desafios podem ser vislumbrados, como a necessidade de, na dimensão epistemológica, se evidenciarem as diferentes correntes teóricas e metodológicas que permeiam o campo assim como suas intersecções, (escolas de pensamento em OC) e suas intersecções; na dimensão aplicada, se evitar o lixo informacional e de se desenvolverem ferramentas mais amigáveis; e, na dimensão cultural, se evitar o preconceito, o proselitismo e as dominações culturais assim como promover compatibilização entre a necessidade de uma comunicação global e o respeito às questões locais.

Por outro lado, interessantes perspectivas se apresentam, como a consolidação da OC como domínio de conhecimento e a configuração de espaços investigativos inter e transdisciplinares no campo (dimensão epistemológica), a promoção da interoperabilidade e de uma recuperação da informação cada vez mais rápida, eficaz e culturalmente significativa (dimensão aplicada) e, na dimensão cultural, o estabelecimento de uma “ética transcultural de mediação, (GARCÍA GUTIÉRREZ, 2002), permeada pela garantia cultural (BEGHTOL, 2002, 2005) e pelo respeito aos domínios de conhecimento (THELLEFSSEN & THELLEFSSEN, 2004).

Tais aspectos nos levam a crer em um futuro promissor para o campo da OC, com a possibilidade de criação de novos capítulos da ISKO (nacionais e regionais), significando novos e diversificados espaços de produção e de interlocução científica, a intensificação e multiplicação das parcerias científicas interinstitucionais e interpaíses, a consolidação desse domínio de conhecimento como espaço investigativo nuclear na Ciência da Informação, e o crescimento (em abrangência e em profundidade) da pesquisa na área, permeada pela necessidade de uma crítica constante e aguçada, inerente aos espaços dinâmicos.

Nesta oportunidade, a ISKO-Brasil promove seu terceiro congresso, sob o tema *Organização do conhecimento do conhecimento de diversidade cultural*, aspecto que encontra lastro na realidade ora vivida pela país e nas preocupações daí decorrentes, e, ainda, lança bases para a 14th ISKO Conference, a realizar-se no Rio de Janeiro, em setembro de 2016, sob o tema *Organização do conhecimento para um mundo sustentável: desafios e perspectivas para o compartilhamento cultural, científico e tecnológico em uma sociedade interconectada*.

O presente congresso de ISKO-Brasil conta, dessa forma, com um conjunto de setenta e cinco trabalhos, entre conferências e comunicações, abordando as três dimensões anteriormente citadas, assim como propiciando um reflexão acerca das perspectivas de pesquisa em OC no Brasil, e sobre a influência de Jean-Claude Gardin na pesquisa em análise documentária no Brasil.

Nesse contexto, há de se registrar o papel do Programa de Pós-graduação em Ciência da Informação da UNESP, como entidade co-promotora e, ainda, o importante apoio recebido da FAPESP, do CNPq, da Capes, da FUNDUNESP e da VUNESP, aspectos que permitiram a realização do evento.

Destaca-se, ainda, presença de quatro resumos que, por haverem sido selecionados como melhores trabalhos, serão publicados, na íntegra, na revista Knowledge Organziation, e, ainda de outros três trabalhos, que receberam o prêmio ISKO-Brasil, a serem publicados na íntegra na revista Brazilian Journal of Information Science – BRAJIS.

Espera-se, com a presente obra, contribuir para a sedimentação científica do campo de OC no Brasil, assim como para a verticalização de suas abordagens, propiciando, ainda, e cada vez mais, a visibilidade e a interlocução científica.

Referências

ANDERSEN, J. Communication technologies and the concept of knowledge organization. **Knowledge Organization,** 2002, 29: 29-39.

BEAK, J.; OLSON, H. A. Analysis of metadata schemas for children's libraries. In: Smiraglia, R. P., ed., **Proceedings from North American Symposium on Knowledge Organization***,, 3*, Toronto, 2011, pp. 13-22. Available: http://iskocus.org/nasko2011-proceedings.php.

BEGHTOL, C. A proposed ethical warrant for global knowledge representation and organization systems. **Journal of Documentation,** London, v. 58, n. 5, p. 507-532, 2002.

BEGHTOL, C. Ethical decision-making for knowledge representation and organization systems for global use. **Journal of the American Society for Information Science and Technology,** New York, v. 56, n. 9, p. 903-912, 2005.

BUCHEL, O.; HILL, L. (2009). Treatment of Georeferencing. in Knowledge Organization Systems: North American Contributions to Integrated Georeferencing. In: Jacob, E. K.; Kwasnik, B. (Eds.). **Proceedings from North American Symposium on Knowledge Organization**, 2009, 2. Syracuse, NY, pp: 47-57. Available: http://iskocus.org/nasko2009-proceedings.php.

BUCKLAND, M. What is a docuemnt? **Journal of the American Society for Information Science,** v.48, n.9, p.804-809, 1997.

BUCKLAND, M. Information as thing. **Journal of the American Society of Information Science,** Washington, v. 42, n. 5, p. 351-360, 1991.

CAMPBELL, D. G. Tensions between language and discourse in north american knowledge organization. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). **Proceedings from North American Symposium on Knowledge Organization,** 2009, 2. Syracuse, NY, pp: 10-16. Available: http://iskocus.org/nasko2009-proceedings.php.

DAHLBERG, I. Knowledge organization: its scopes and possibilities. **Knowledge organization**, Frankfurt, n. 20, p. 211-222, 1993

DAHLBERG, I. Current trends in Knowledge Organization. In: GARCIA MARCO, F. J. **Organización del conocimiento en sistemas de información y documentación** 1. Actas del I Encuentro de ISKO-España, 1993, Madrid. Zaragoza: Librería General, 1995. p. 7-25.

DOUSA, T.M. Classical Pragmatism and its Varieties: On a Pluriform Metatheoretical Perspective for Knowledge Organization. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). **Proceedings from North American Symposium on Knowledge Organization,** 2. Syracuse, NY, pp:1-9. Available: http://iskocus.org/nasko2009-proceedings.php.

EDWARDS, C. (2011). Soviet classifications and situated knowledges. In SMIRAGLIA, Richard P., ed. **Proceedings from North American Symposium on Knowledge Organization**, 2011, 3. Toronto, pp. 13-22. Available: http://iskocus.org/nasko2011-proceedings.php.

FORHMANN, B. **Deflating information:** from science studies to documentation. Toronto: University of Toronto Press, 2004.

FROHMANN, B. Revisting “What is a docuemnt”. **Journal of Documentation,** v.6, n.2, p.291-303, 2009.

GARCÍA GUTIÉRREZ, A. Knowledge organization from a “culture of the border”: towards a transcultural ethics of mediation. In: LÓPEZ-HUERTAS, M. J. (Ed.). Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries. Würzburg: ERGON-Verlag, 2002. p. 516-522.

GNOLI, C. Ten long-term research questions in knowledge oragnzation. Knowledge organization, v.35, n.2/3, p.137-149, 2008.

GNOLI, C. Ontological foundations in KO: the theory of integrative levels applied in citation order. **Scire.** 2011, 17(1): 29-34.

GREEN, R.; FALLGREN, N. (2007) Anticipating new media: A faceted classification of material types. In Tennis, J.T., Ed. **Proceedings North American Symposium on Knowledge Organization,** 2007, *1.* Toronto, pp. 87-99. Available: http://iskocus.org/nasko2007-proceedings.php.

GREEN,R. See-also relationships in the Dewey Decimal Classification. In Smiraglia, Richard P., ed. **Proceedings from North American Symposium on Knowledge Organization,** 2011, 3. Toronto, pp. 41-51. Available: http://iskocus.org/nasko2011-proceedings.php.
HJØRLAND, B. Deliberate bias in knowledge organization? In: ARSENAULT, C.; TENNIS, J. T. (Org.). **Culture and identity in knowledge organization.** Würzburg: Ergon Verlag, 2008. p. 256-261. (Advances in knowledge organization, 11).

HJORLAND, B. Fundamentals of knowledge organization. **Knowledge Organization**, v. 30, n. 2, p. 87-111, 2003.

HJØRLAND, B. Fundamentals of knowledge organization. **Knowledge Organization,** Würzburg, v. 30, n. 2, p. 87-111, 2003.

HJORLAND, B. Nine principles of knowledge organization. In: ALBRECHTSEN, H.; ORNAGER, S. (ed.) **Knowledge organization and quality management**. Frankfurt/main: Indeks, 1994. p. 91-100. (Advances in Knowledge Organizatio, v. 4).

HJORLAND, B. What is knowledge organization (KO). **Knowledge organization,** Frankfurt, v. 35, n. 2 –3, p.86-101, 2008.

HOFFMAN, G.L. (2009) Applying the User-Centered Paradigm to Cataloging Standards in Theory and Practice: Problems and Prospects. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 27-34. Available: http://iskocus.org/nasko2009-proceedings.php.

HUDON, M. Accessing documents and information in a world without frontiers. **The Indexer**, London, v. 21, n. 4, p. 156-159, 1999.

HUDON, M. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. **Knowledge Organization,** Würzburg, v. 24, n. 2, p. 84-91, 1997.

ISKO´s mission. Disponíovel em: http://www.isko.org/about.html. Acesso em: 02.09.2015.

KIPP, M. E. I. (2007). Tagging for health information organisation and retrieval. In Tennis, J. T., Ed. **Proceedings of the North American Symposium on Knowledge Organization,** *1.* Toronto, pp. 63-74. Available: http://iskocus.org/nasko2007-proceedings.php.

LÓPEZ-HUERTAS, M. J. Some current research questions in the field of knowledge organization. **Knowledge Organization,** Würzburg, v. 35, n. 2/3, p. 113-136, 2008.

MILANI, S. O.; GUIMARÃES, J. A.C. (2011). Biases in knowledge representation: an analysis of the feminine domain in Brazilian indexing languages. In Smiraglia, Richard P., ed. Proceedings from North **American Symposium on Knowledge Organization**, 3. Toronto, pp. 94-104. Available: http://iskocus.org/nasko2011-proceedings.php.

OLSON, H. A. **The power to name:** locating the limits of subject representation in libraries. Dordrecht: Kluwer Academic Publishers, 2002.

SMIRAGLIA, R. P. Modulation and Specialization in North American Knowledge Organization: Visualizing Pioneers In: In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). **Proceedings from North American Symposium on Knowledge Organization,** 2009, 2. Syracuse, NY, pp: 35-46. Available: http://iskocus.org/nasko2009-proceedings.php.

SMIRAGLIA, R.P. Epistemology of domain analysis. In **Cultural frames of knowledge**. Würzburg: Ergon-Verlag, 2012, pp. 111-24

TENNIS, J. T. Epistemology, theory and methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework. **Knowledge organization**, Frankfurt, v. 35, n. 2-3, p. 102-112, 2008.

TENNIS, J.T. "Epistemology, Theory, and Methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework.", 2008. In: **Knowledge Organization. 35(2/3): 102-112.**

THELLEFSEN, T. L; THELLEFSEN, M. M. Pragmatic semiotics and knowledge organization. **Knowledge Organization**, Würzburg, v. 31, n. 3, p. 177-187, 2004.

ZINS, C. (2004). Knowledge organization: an epistemological perspective. **Knowledge organization** 31: 49-54.

A Dimensão Epistemológica da Organização e Representação do Conhecimento

Epistemic, Ontic, Axiologic, and Praxic Constructs in Knowledge Organization Research

Joseph T. Tennis
University of Washington
jtennis@uw.edu

Introduction

The work of knowledge organization requires a particular set of tools. For instance we need standards of content description like Anglo-American Cataloging Rules Edition 2, Resource Description and Access (RDA), Cataloging Cultural Objects, and Describing Archives: A Content Standard. When we intellectualize the process of knowledge organization – that is when we do basic theoretical research in knowledge organization we need another set of tools. For this latter exercise we need *constructs*. Constructs are ideas with many conceptual elements, largely considered subjective. They allow us to be inventive as well as allow us to see a particular point of view in knowledge organization. For example, Patrick Wilson's ideas of exploitative control and descriptive control, or S. R. Ranganathan's fundamental categories are constructs. They allow us to identify functional requirements or operationalizations of functional requirements, or at least come close to them for our systems and schemes. They also allow us to carry out meaningful evaluation.

What is even more interesting, from a research point of view, is that constructs once offered to the community can be contested and reinterpreted and this has an affect on how we view knowledge organization systems and processes. Fundamental categories are again a good example in that some members of the Classification Research Group (CRG) argued against Ranganathan's point of view. The CRG posited more fundamental categories than Ranganathan's five, Personality, Matter, Energy, Space, and Time (Ranganathan, 1967). The CRG neededsignificantly more

fundamental categories for their work.[1] And these are just two voices in this space we can also consider the fundamental categories of Johannes Kaiser (1911), Shera and Egan, Barbara Kyle (Vickery, 1960), and Eric de Grolier (1962). We can also reference contemporary work that continues comparison and analysis of fundamental categories (e.g., Dousa, 2011).

In all these cases we are discussing a construct. The fundamental category is not discovered; it is constructed by a classificationist. This is done because it is useful in engaging in the act of classification. And while we are accustomed to using constructs or debating their merit in one knowledge organization activity or another, we have not analyzed their structure, nor have we created a typology. In an effort to probe the epistemological dimension of knowledge organization, we think it would be a fruitful exercise to do this.This is because we might benefit from clarity around not only our terminology, but the manner in which we talk about our terminology. We are all creative workers examining what is available to us, but doing so through particular lenses (constructs) identifying particular constructs. And by knowing these and being able to refer to these we would consider a core competency for knowledge organization researchers.

In the next section we lay out our preliminary typology of constructs and in so doing identify some attributes of the constructs under investigation. Of course we cannot, in the space of this paper, describe all constructs used in knowledge organization. However, this is a start to a conversation that we hope improves our understanding of the philosophical commitments we make in the practice of knowledge organization research.

Constructs and their Kinds

It is possible for us separate out constructs along a variety of axes. We can consider their philosophical nature, and ask whether they about things in knowledge organization or about how we know about some aspect of knowledge organization.

[1] For example, in the classification of science Brian Vickery proposes: Substance (product), organ, constituent, structure, shape, property, object of action (patient, raw material), action, operation, process, agent, space, and time. See Vickery (1960) p. 23. And see also Austin (1968), Foskett (1974), Mills and Broughton (1977).

The former is an *ontic construct* the latter is an *epistemic construct.* A third category of philosophical construct can be identified as well. *Axiologic constructs* are those that focus on worth – both aesthetic and ethical aspects of knowledge organization. One final construct we consider here is the *praxic construct*. These are constructs that deal with particular actions and activities in knowledge organization. In the following section we provide examples and describe further the attributes of these constructs.

Ontic Constructs: What there is in the world of knowledge organization

As mentioned above, fundamental categories are constructs we use in order to create analytico-synthetic classification (*vide* Ranganathan, 1967). Fundamental categories are *things* we construct in order to do the work at hand. What is interesting about ontic constructs is how researchers agree or disagree with their definitions and boundaries. One such ontic construct is the *subject,* which we have represented as a real-word object that can be faithfully represented in our systems. However the nature of subjects is, in the contemporary literature, often discussed from a pragmatic stance – that of an informed judgment, which has a sometimes-tense relationship to realist ontologies. For example, there is a fundamental disagreement in the literature about the nature of subjects (Kučerová, 2014).By acknowledging these characteristics, we can begin to refine our assumptions and our knowledge of subjects as mobilized in our information systems.

Ontic constructs form the largest set of constructs in knowledge organization literuatre. We have for the past one-hundred and fifty years or so, made *things* that exist in the universe of knowledge organization. The other constructs, while not immediately recognizable as things are just as valuable, because they make up a different kind of construct. The work of the remaining three constructs is not about what *is* in the universe of knowledge organization, but rather what in the universe of knowledge *has worth, allows us to know,* and *what we do.* We will first talk about what in the universe of knowledge organization has worth.

Axiologic Constructs

With the rising concern about the ethics of knowledge organization systems, we have seen a rise of axiologic constructs. These constructs about the value of knowledge organization systems, both ethical and aesthetic, interrogate the worth of our work in knowledge organization. If we are adding value, how do we assess that

value? Classification as rhetorical device, a position advanced by Melanie Feinberg (Feinberg 2011), is one way to look at knowledge organization systems in an aesthetic way. The value presented here is in the classification scheme's ability to persuade me, the user (or perhaps reader) of the scheme of the author's position.

There are also ethical constructs that have surfaced in our contemplation about the potential harm our systems might cause to our users.This line of research is indebted to Sanford Berman (1971) and continues today with contemporary scholarship, with Melissa Adler (2013), Fabio Pino (2010), Melody Fox (2012) following on substantial contributions by Hope Olson(2002) Jonathan Furner (2008), and Clare Beghtol (2002).

Constructs in this space are a bit different than the ontic constructs above. In many cases we have to do some creative naming in order to isolate them from the texts. For example, Guimarães et al., identify ethical values and ethical problems in knowledge organization, while acknowledging a wider concern in ethical issues in contemporary global society (Guimarães et al., 2008). In order to make ethical values and ethical problems into constructs, we might need to compare these lists with others and squarely define what we mean by values and problems. Would Guimarães et al., agree with other thinkers in the field as to whether something is an ethical value or problem for knowledge organization? We would only know this if we had a definition and extension of the construct that can persist through time and across contexts. If there is disagreement among researchers in knowledge organization, then we again see the fruitfulness of constructs – by being explicit about what we are constructing we can come to clarity about the substance of our discussions and, by extension, the way we see knowledge organization systems can be improved.

The bottom line with regard to *axiologic constructs* is that we must be able to articulate value and its correllaries, ethics and aesthetics where appropriate in knowledge organization systems. We can do this by understanding of the ideas present in the practice and reflection of knowledge organization as ideas with conceptual elements – i.e., constructs.

Epistemic Constructs

Ontic and axiologic constructs say what is and what has value in knowledge organization. However, a core concern in knowledge organization is epistemology and epistemological issues. These issues at their most basic ask *how do we know* some

aspect of knowledge or knowledge organization. In an applied way, knowledge organization has also crafted epistemology and epistemic concerns, to be the purview of theories of knowledge. That is, it is not only concerned with how we know, but clarifying our assumptions verses our intentions with how we know. Further, investigating nad manufacturing epistemic constructs can give rise other constructs, particularly ontic constructs

David Blair is one fundamental thinker in this space when he argues for a particular epistemic stance on language and information retrieval (Blair, 1990; 2005).In the case of Blair, we have the epistemic construct of *language as speech act* where we claim to know the ontic construct (language) in a particular way (as a speech act). This is a refinement of the assumption that we are dealing with the pragmatic view of the universe of language. The relationship between the universe of language and the universe of knowledge is a core concern for knowledge organization, and by making explicit what that relationship is, if they are indeed separate and separable constructs is the business of epistemic constructs we deploy in knowledge organization research.

For example in reference theory, we construct an epistemic construct of the concept triangle with its three parts: symbol, thought, referent (Ogden and Richards, 1923). This is an epistemic construct because it points to how we should know the relationship between language and knowledge.

Named different schools of thought can be seen as epistemic constructs too. So rationalist, empiricist (Hjørland, 2005), analogic (Lee, 2012), indigenous imagining (Duarte and Belarde-Lewis, 2015), are all constructs that allow us views into what ontic constructs are available to us, and informs our axiologic constructs. For example, the concept of harmony might not be important to rationalist or empiricist constructs, but that cannot be said of analogic or indigenous imagining – where what is known is couched in a deep relationship to a ethos of harmonious being in relation to the cosmos. Again we can mention Thomas Dousa because work fits here as well. He is very concerned with rendering clear the meaning behind labels. For example he has probed what it means to say one is taking a pragmatist's stance in relation to knowledge organization (Dousa, 2010).

The final example presented here is the concept of naïve classification and professional classification (Beghtol, 2003). What we know as classification and how we conceptualize its place in the universe of knowledge organization research is an epistemic construct. This is born out by the debate over what was actually under

discussion in Beghtol's article (Nicholaisen and Hjørland, 2004; Jacob, 2010). We have a tradition of naming these after the authors, and that seems reasonable to continue if not further encourage.

Praxic Constructs

The final constructs are those that deal with the act and action of knowledge organization. Indexing, classification, domain analysis, cataloguing are all actions and each of them can be defined and described in a variety of ways.This makes them constructs. Further, any subdivision of these acts can be considered constructs, named and therefore recognized as objects of investigation. Jens-Erik Mai has described a three step indexing process (Mai, 2003). He subdivided subject indexing into (1) document analysis, (2) subject description process, (3) subject analysis process (Mai, 2003 p. 595).Each of these is an action and he has offered them as praxic constructs whereby we can probe further (in agreement or disagreement with others) into how indexing occurs.

Interestingly we have thought tagging on social media websites to be similar to indexing (Tennis, 2006; Kip 2014). The assumption here seems to be a bit at the intersection between epistemic and praxic constructs because we are assuming one way of identifying indexing

Next Steps

If constructs are useful for knowledge organization research to consider we have a number of next steps we can take. The first is to identify the population of constructs that exist in the knowledge organization literature. The next is to figure out which are the same constructs under different names and which are different constructs under the same name. And finally we need to consider which are core to our present concerns in knowledge organization, which are orthogonal or corollary – which are not core, but play important roles at times in our discussions of knowledge organization research, and which are outmoded, perhaps too rooted to extinct technology or outdated functional requirements. An example of core constructs might be facets, ethical values, pragmatism, and indexing all mentioned above. An orthogonal construct might be indigenous imagining where we are a concerned with involving Native American people in designing knowledge organization systemsor a separate construct for the action of tagging in the context of Twitter. Outmoded

constructs could always come back based on their value or utility, but we do not as a matter of routine talk about *abstract classification* the way S. R. Ranganathan did in the 1960s (Ranganathan, 1967). It is possiblethat we might talk about abstract classification sometime in the future, but we are no longer actively working on this in this community.

Conclusion

This paper has offered the idea of a family of constructs in knowledge organization research. I have done this in an effort to probe the epistemological dimension to knowledge organization. What we know and how we talk about what we know in knowledge organization is deeply rooted in two camps: philosophy and practice. We have, through the natural course of our work, naively created constructs to help us work with our universe of knowledge organization. Classes, facets, ethical problems, empiricism, the act of cataloguing are all constructs borrowed from philosophy or practice and bound together in a particular way we use language in knowledge organization. They are ours as much as they are anyone else's. However, we must appreciate our particular needs in naming these in our literature. We would argue that we need these to be constructs. We need to deploy these to create, maintain, and evaluate knowledge organization systems. We have a particular use for these ideas that we have seen in the literature of philosophy and the everyday life of practice.

Knowledge organization systems are the results of a deeply philosophical and imminently practical activity of making the world's recorded knowledge available to users. We have approached this problem by creating our systems out of our understanding of knowledge and users, which includes language, technology, context, and values. By probing deeper into constructs as part of the epistemological dimension we move toward a systematic understanding of how we have crafted the relationship between philosophy and practice and how we have created an inventory of our contributions to helping users access the universe of knowledge.

References

Adler, M. A. (2013). The ALA Task Force on Gay Liberation: Effecting Change in Naming and Classification of GLBTQ Subjects. Advances In Classification Research Online, 23(1), pp. 1-4

Austin, D. (1968). *Fields, categories and general systems theory: Report to the Classification Research Group*. (London: BNB).

Berman, S. (1971).*Prejudices and Antipathies* (Metuchen, NJ: Scarecrow Press).

Beghtol, C. (2002). “A proposed ethical warrant for global knowledge representation and organization systems.” In *Journal of Documentation*58(5): 507-532.

Beghtol, C. (2003). “Classification for Information Retrieval and Classification for Knowledge Discovery: Relationships between “Professional” and “Naïve” Classifications.” In *Knowledge Organization* 30(2): 64-73

Blair, D. C. (1990). *Language and Representation in Information Retrieval.* (Elsevier).

Blair, D. C. (2005). Information retrieval and the philosophy of language. In *Annual Review of Information Science and Technology*.

Dousa, T. M. (2010). “Classical Pragmatism and its Varieties: On a Pluriform Metatheoretical Perspective for Knowledge Organization.” *Advances in Knowledge Organization Research*

Dousa, T. M. 2011. Concretes, countries, and processes in Julius O. Kaiser's Theory of Systematic Indexing: A case studyin the definition of general categories. In Smiraglia, Richard P., ed. *Proceedings from North American Symposium on KnowledgeOrganization*, Vol. 3. Toronto, Canada, pp. 160-173.

Duarte, M. E. and Belarde-Lewis, M. (2015). “Imagining: Creating Spaces for Indigenous Ontologies.” In *Cataloging and Classification Quarterly* 53(5-6): 677-702.

Feinberg, M. (2011). Feinberg. Compiler to author: a process for designing rhetorically aware document collections. Journal of the American Society for Information Science and Technology 62(9): 1784-1796.

Foskett, D. J. (1974). *Classification and indexing in the social sciences*. (London: Butterworths).

Fox, M.J. & Olson, H.A. (2012). Feminist epistemologies and knowledge organization. In Hur-li Lee and Richard Smiraglia (eds.), *Cultural frames of knowledge*. Wurzburg: Ergon Verlag. 79-98.

Furner, J. (2008). Dewey deracialized: A critical race-theoretic perspective. In *Knowledge Organization* 43(3): 144-168.

Grolier, E. de. (1962). *A study of general categories applicable to classification and coding in documentation.*(Paris: UNESCO).

Guimarães, J. A. C., Fernández-Molina, J. C., Pinho, F. A., Milani, S. O. (2008). Ethics in the Knowledge Organization Environment: An Overview of Values and Problems in the LIS Literature.” In *Advances in Knowledge Organization* :361-366.

Hjørland, B. (2005). “Empiricism, rationalism and positivism in library and information science.” In *Journal of Documentation* 61(1): 130-155.

Jacob, E. K. (2010). “Proposal for a Classification of Classifications built on Beghtol’s Distinction between “Naïve Classification” and “Professional Classification”” In *Knowledge Organization* 37(2): 111-120.

Kaiser, J. (1911). *Systematic Indexing.* (Isaac Pitman & Sons).

Kipp, M. (2014). “Understanding the retrieval effectiveness of collaborative tags and author keywords in different retrieval environments: An experimental study on medical collections.” In *Journal of the Association for Information Science and Technology* 65(3): 483-500.

Kučerová, H. (2014). “What Do we Analyze during Subject Analysis of Documents? The Concept of Aboutness in Knowledge Organization.” In *Revue of Librarianship* 25(1): 38-57.

Lee, H-L. (2012). “Epistemic Foundation of Bibliographic Classification in Early China: A Ru Classicist Perspective.” In *Journal of Documentation* 68(3): 378-401.

Mai, J-E. (2003). “Semiotics and indexing: An analysis of the subject indexing process. In *Journal of Documentation* 57(5): 591-622.

Mills, John, and Vanda Broughton. (1977). *Bliss bibliographic classification. Second edition*. (London: Butterworths).

Nicholiasen, J. and Hjørland, B. (2004). “A rejoinder to Beghtol.” In *Knowledge Organization* 31: 199-201.

Ogden, C. K. and Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism.* (Trench, Trubner & Co)

Olson, H. A. (2002). *Power to name.* (Springer).

Pinho, F. A. (2010). Aspectos éticos em representação do conhecimento em temáticas relativas à homossexualidade masculina: uma análise da precisão em linguagens de indexação brasileiras. Dissertation for UNESP.

Ranganathan, S. R. (1967). *Prolegomena to Library Classification.* 3rd Edition. (Bombay: Asia Publishing House).

Tennis, J. T. (2006). “Social Tagging and the Next Steps for Indexing: Fordist Reflexivity and Intertextuality.” In *Advances in Classification Research*. vol. 17. (Austin, Texas). Available: http://dlist.sir.arizona.edu/1726/

Vickery, B. C. (1960). *Faceted classification: A guide to construction and use of special schemes.*

Abordagens e paradigmas na organização do conhecimento

Leila Cristina Weiss
Universidade Federal de Santa Catarina
leilacw@gmail.com

Marisa Bräscher
Universidade Federal de Santa Catarina
marisa.brascher@gmail.com

1 Introdução

Existem diferentes propostas de classificação das correntes teórico-metodológicas existentes no campo de Organização do Conhecimento (OC) que, de modo geral, são frutos de discussões epistemológicas sobre o campo, que é interdisciplinar e recebe contribuições de áreas como a Ciência da Informação, Filosofia, Linguística, Semântica, Ciências da Computação, dentre outras.

Uma das propostas de classificação é a de Hjørland (2007) e se desenvolveu a partir da análise dos paradigmas em semântica, devido à forte influência dessa área sobre o campo de OC. Sob essa perspectiva, distinguem-se as abordagens pragmática e positivista em OC.

Pragmático, positivista, operacionalista, referencial, instrumental, empirista, racionalista e realista são alguns nomes de posturas epistemológicas, segundo Tennis (2008). A proposta de classificação das correntes teórico-metodológicas presentes na OC desenvolvida por Gnoli (2008), no entanto, não se refere a nenhuma destas posturas epistemológicas, mas sim à abordagem ontológica e à abordagem epistemológica, que o autor considera serem as duas abordagens básicas na OC.

Devido à importância das propostas de Hjørland (2007) e Gnoli (2008), refletidas nas recentes pesquisas e discussões na área de OC, neste trabalho analisa-se comparativamente essas duas correntes teórico-metodológicas na OC. Hjørland (2007) distingue os paradigmas na OC em pragmático e positivista, e Gnoli (2008), distingue as abordagens na OC em ontológica e epistemológica. Descrevem-se cada

uma das propostas e verificam-se as semelhanças, diferenças e complementariedades entre elas.

2 Paradigmas pragmático e positivista na OC

Iniciamos a discussão sobre a proposta de Hjørland (2007) para a classificação das correntes teórico-metodológicas nas pesquisas em OC a partir de um trabalho anterior do mesmo autor, publicado em 2003, no qual se descreve a presença das posturas epistemológicas empirista, racionalista, historicista e pragmática na OC.

Hjørland (2003) aponta que os métodos de OC estão relacionados com pressupostos epistemológicos, dos quais destaca o empirismo, o racionalismo, o historicismo e o pragmatismo. De acordo com o autor, na OC alguns artigos e sistemas apoiam-se principalmente em generalizações empíricas, como por exemplo os sistemas com base na medida de frequência de palavras. Outros artigos e sistemas são, em sua maioria, adotam regras racionais e deduções (embora muitas vezes ignorando questões empíricas). Nesse caso o autor considera que os sistemas analítico-facetados na tradição de Ranganathan fornecem os melhores exemplos. Um terceiro tipo de sistema tem por base o estudo da evolução de campos do conhecimento (historicismo). Em certo grau os sistemas com bases disciplinares, como a Classificação Decimal de Dewey (CDD), estariam em conformidade com esse ideal. Mas, por outro lado, a atualização desse sistema é mais influenciada pelos princípios da tradição analítico-facetada (como discutido por Miksa, 1998), e, portanto, em certo sentido a CDD também teria uma base empírica. Já os estudos e sistemas com base no pragmatismo são desenvolvidos com base na análise de objetivos, valores e consequências. O autor ressalta que esse paradigma na OC se insere na vertente realista, pois as construções sociais produzidas somente a partir de interesses pessoais ou dos pesquisadores da área, quando confrontadas com a realidade, podem se mostrar incoerentes. Dessa maneira, o método pragmático não se opõe aos aspectos do empirismo, racionalismo e historicismo. No entanto, evidências isoladas não são suficientes para apoiar os processos de OC e as evidências necessárias estão relacionadas aos objetivos e às atividades humanas. (HJØRLAND, 2003).

Nessa direção, Tennis (2008) argumenta que a epistemologia envolve uma decisão sobre que conhecimento é ‘válido’ para a OC e o que pode ser considerado

como fonte de evidência aceitável sobre esse conhecimento. Portanto, “a epistemologia em OC resulta numa posição epistêmica que esboça uma reivindicação de conhecimento” (TENNIS, 2008, p.105)

Esses pressupostos epistemológicos sobre a linguagem são discutidos por Hjørland (2007), que considera a semântica uma área de extrema importância para a OC. De acordo com o autor, abordagens distintas para OC implicam diferentes pontos de vista sobre semântica.

Peregrin (2004) destaca dois paradigmas dominantes em semântica, o pragmático e o positivista. No paradigma positivista a linguagem é retratada como um meio de representar o mundo e a realidade. No paradigma pragmático, a linguagem é vista como um meio de interação e o significado de uma expressão não é a entidade representada por ela, mas sim a sua função na interação. Com base em Peregrin (2004), Hjørland (2007) distingue os paradigmas em OC em pragmático e positivista, diferentemente de como havia definido os paradigmas em seu trabalho de 2003.

Francelin e Kobashi (2011), de modo semelhante, distinguem as abordagens sobre o conceito na OC em positivista e pragmática. Em seu estudo, com base na literatura brasileira publicada em artigos científicos, esses pesquisadores concluem que “as linhas de força teóricas estão calcadas, fundamentalmente, na lógica aristotélica, mais bem representada pelas Categorias, e pelo princípio de fixação da linguagem ou conceitos por meio de juízos e proposições verdadeiras, características básicas da filosofia analítica e do positivismo lógico” (FRANCELIN; KOBASHI, 2011, p.221).

Por outro lado, ainda segundo os autores supracitados, é possível encontrar fundamentação pragmática sobre o uso da linguagem e dos conceitos, sobretudo nos artigos que “analisam o conceito na perspectiva da filosofia da linguagem, da semiótica, da linguística, da teoria comunicativa da terminologia e da socioterminologia”. (FRANCELIN; KOBASHI, 2011, p.222).

Tanto Hjørland (2007) quanto Francelin e Kobashi (2011) apontam que no geral as pesquisas em OC não indicam de forma explícita em qual teoria ou corrente filosófica se baseiam. Mas ainda assim indicam forte inclinação para o positivismo, e Francelin e Kobashi (2011), afirmam também que muitas vezes, utilizam fundamentos dessa corrente teórica, como por exemplo a lógica formal e as categorias e lógica aristotélicas, de maneira acrítica. Hjørland (2009) trata a teoria do conceito sob a ótica

epistemológica e considera as visões pragmática e historicista como as mais produtivas para o desenvolvimento da OC.

Assim, para os fins da análise desse trabalho, a proposta de classificação considerada foi a de Hjørland (2007), que distingue os paradigmas em OC apenas em pragmático e positivista. De modo geral, pode-se considerar uma adaptação de sua proposta inicial publicada em 2003, pois, em sua proposta de 2007 os paradigmas racionalista e empirista se enquadrariam no paradigma positivista, e os paradigmas historicista e pragmático, se enquadrariam no pragmático.

3 Abordagem ontológica e epistemológica na OC

Gnoli (2008) inicia sua descrição das abordagens ontológica e epistemológica na OC a partir de considerações sobre a Ontologia e a Epistemologia, sendo que ambas são subcampos da filosofia. Nesse contexto, afirma que a Ontologia

> [...] diz respeito à natureza das coisas conhecidas, especialmente em termos de categorias gerais a que pertençam. Questões como a subdivisão de uma classe em tipos e partes, ou o reconhecimento de que determinado conceito consiste num processo ou numa entidade estática, são ontológicas. Epistemologia, por outro lado, diz respeito a como os seres humanos conhecem o mundo por meio dos órgãos dos sentidos, e como eles processam o conhecimento de acordo com categorias tanto inatas quanto e culturalmente adquiridas (GNOLI, 2008, p.139).

Na visão desse mesmo autor, “o conhecimento é tanto epistemológico quanto ontológico, pois envolve a percepção humana e também se refere a objetos do mundo real que, por sua vez possuem certa estrutura intrínseca.” (GNOLI, 2008, p.139). O conhecimento documentado, a informação, carrega essas características. Redón Rojas (1996) considera que a informação é construída e não simplesmente inventada (subjetiva) ou descoberta (objetiva). A informação não é uma simples invenção sua elaboração apoia-se em certa matéria prima (elementos objetivos que se encontram nas coisas); e tampouco é algo pronto e objetivo que só precisa ser descoberto, pois é necessária à ação do homem. Os homens transformam e são transformados pelo mundo informacional em um processo de evolução constante. Nesse conceito de informação, a dialética do sujeito com o mundo que o rodeia é levada em consideração.

Ainda na visão de Gnoli (2008), na abordagem ontológica o foco recai sobre o mundo e a suposta realidade objetiva, enquanto que na abordagem epistemológica o foco recai sobre o sujeito e o contexto social, conforme exemplifica:

> As principais classes de Dewey seguem uma sequência epistemológica, como sugere Francis Bacon, são listadas conforme a sequência em que são produzidas pelo intelecto humano: razão, imaginação e memória; as principais classes *da CDU* são também epistemológicas, uma vez que são derivados de Dewey. Outros sistemas, como a *Bliss Classification*, a *Broad System of Ordering*, e a I*nformation Coding Classification*, embasam a sequência de suas principais classes naquela que é supostamente natural, de acordo com a especificidade e complexidade dos objetos conhecidos, são, portanto, principalmente ontológicos (GNOLI, 2008, p. 139).

Prosseguindo com a distinção entre essas duas abordagens na OC, Gnoli (2008), ainda cita Dahlberg (1974; 1978) e Robert Poli, que fazem referência ao filósofo Nicolai Hartmann, como representantes da abordagem ontológica e Hjørland e Albrechtsen (1995) como representantes da abordagem epistemológica. Os pesquisadores em OC que desenvolvem suas pesquisas sob a perspectiva da abordagem epistemológica, segundo Gnoli (2008, p.19), "muitas vezes tomam exemplos a partir da linguagem e seu relativismo cultural, e citam filósofos como Wittgenstein e os pragmáticos americanos".

O autor não defende uma ou outra abordagem como sendo melhor ou mais apropriada para a OC, mas sim que ambas devem se reconciliar e serem trabalhadas juntas. Os problemas que podem ocorrer quando a abordagem ontológica predomina em uma pesquisa em OC ou no desenvolvimento de um SOC não chegam a ser tratados pelo autor, que apresenta, no entanto, algumas considerações sobre os problemas da predominância da abordagem epistemológica.

> Os defensores do relativismo cultural insistem que é melhor se concentrar em esquemas de domínios específicos, e ter ciência de suas bases epistemológicas. Pois qualquer esquema será influenciado pelo contexto cultural no qual foi concebido (Hjørland 2004). No entanto, a referência a um regime geral é necessária, mesmo quando a indexação é de literatura específica. (DJ Foskett 1991). Se realmente queremos permitir a interoperabilidade entre diferentes esquemas e a pesquisa interdisciplinar, vamos sempre precisar de algum regime geral, pelo menos como um dispositivo de comutação entre os sistemas sustentados em diferentes epistemologias. Assim, a necessidade de investigação ontológica em OC está longe de ser um tema obsoleto. (GNOLI, 2008. p. 140).

Essa afirmação, em nosso entendimento, justifica a importância de as duas abordagens serem trabalhadas conjuntamente na OC.

4 Considerações finais

De modo geral, estudos epistemológicos, como os desenvolvidos por Hjorland e Gnoli, se referem às correntes teórico-metodológicas de uma área como paradigmas, ainda que tomem por base critérios de classificação distintos. Podemos dizer que as propostas de Hjørland (2007) e Gnoli (2008) para a classificação das correntes teórico-metodológicas na OC foram desenvolvidas a partir de critérios classificatórios distintos. A primeira distingue os paradigmas na OC em pragmático e positivista e adota como critério as abordagens epistemológicas. Enquanto que a segunda distingue as abordagens em ontológica e epistemológica e adota a forma de ver o objeto de estudo da OC como critério. Consideramos, portanto, que essa parece ser a principal diferença entre as duas propostas, o critério de classificação.

Cada postura epistemológica possui critérios de validação do conhecimento próprios e utiliza métodos de pesquisas que visam atender a esses critérios. Isso faz com que um mesmo objeto de pesquisa seja visto de forma distinta, de acordo com a postura epistemológica adotada na pesquisa. Com uma postura epistemológica pragmática se vê o objeto de estudo da OC como epistemológico e com uma postura epistemológica positivista se vê o objeto de estudo da OC como ontológico. Assim, apesar dessa diferença, verifica-se que essas duas propostas de classificação das correntes teórico-metodológicas na OC são complementares. Por exemplo, a Classificação Decimal de Dewey (CDD) é classificada por Hjørland (2003) como integrante do paradigma historicista (pragmático) na OC, e Gnoli (2008), utiliza a CDD e a Classificação Decimal Universal (CDU) como exemplos da abordagem epistemológica na OC. Hjørland (2003) atribui os sistemas analítico-facetados da tradição de Ranganathan como integrantes do paradigma racionalista (positivismo), pois estão apoiados em regras racionais e deduções. Ranganathan desenvolveu seu sistema de classificação com base nos trabalhos de Henry Evelyn Bliss, a *Bliss Classification*, bem como a *Broad System of Ordering*(BSO) e a *Information Codig Classification* (ICC), são consideradas por Gnoli (2008) como integrantes da abordagem ontológica.

Além dessas semelhanças, verifica-se que tanto Birger Hjørland quanto Claudio Gnoli, consideram importante integrar as abordagens na OC. Hjørland, apesar de não enfatizar este aspecto em seus trabalhos, assim como ocorre em Gnoli (2008), deixa claro que a abordagem pragmática não se opõe às demais posturas epistemológicas, apenas considera que as evidências isoladas não são suficientes e que os critérios de verdade estão ligados aos objetivos e às atividades humanas. (HJØRLAND, 2003, p. 106). Nenhum dos dois autores propõe como se daria essa complementariedade e, talvez, as pesquisas na área possam avançar nessa direção, e, ainda, na consolidação de metodologias que integrem essas visões.

Referências

FRANCELIN, M. M; KOBASHI, N. Concepções sobre o conceito na organização da informação e do conhecimento. **Ciência da Informação**, Brasília, v. 40 n. 2, 2011.

GNOLI, C. Ten Long-Term Research Questions in Knowledge Organization. **Knowl. Org.**, v.35, n. 2, p.137-149, 2008.
HJØRLAND, B. Concept theory. **JASIST**, v.60, n.8, p.1519-1536, 2009.

HJØRLAND, B. Fundamentals of Knowledge Organization. **Knowl. Org.**, v.30, n.2, p.87-111, 2003.

HJØRLAND, B. Semantic and Knowledge organization. **ARIST**, v.41, p. 367-405,2007.

PEREGRIN, J. Pragmatism und Semantik. In: Fuhrmann, A.; OLSSON, E. J. **Pragmatisch denken** . p. 89-108. Frankfurt am Main, Germany: Ontos. 2004.

RENDÓN ROJAS, M. A. Hacia um nuevo paradigma em bibliotecologia. **Transinformação**, Campinas, v. 8, n. 3, p. 17-31, set./dez. 1996.

TENNIS, J. T. Epistemology, Theory, and Methodology in Knowledge Organization: Toward a Classification, Metatheory, and Research Framework. **Knowl. Org.**, v. 5, n.2, p. 102-112. 2008.

A dimensão histórico-epistemológica da Organização do Conhecimento: contributos da taxonomia Gesneriana, Séc. XVI

Andre Vieira de Freitas Araujo
Universidade Federal do Rio de Janeiro
armarius.araujo@gmail.com
armarius.araujo@gmail.com

Giulia Crippa
Universidade de São Paulo
giuliac@ffclrp.usp.br

1 Introdução

A dimensão histórico-epistemológica da organização do conhecimento (OC) é, evidentemente, necessária para compreensão de sua natureza e de seus fundamentos.

São inúmeros os marcos históricos que constituem o campo. Seu resgate possibilita a constante revisão de aspectos práticos e teóricos e a formulação de novos entendimentos e respostas para antigos e novos problemas que delineiam a paisagem da informação e do conhecimento.

Smiraglia (2013) afirma que se a OC é a ciência da ordem do conhecimento, então seus cientistas devem se preocupar em primeiro lugar com a questão do que é, exatamente, o conhecimento.

Iluminando esta perspectiva, o resgate de fontes documentais sobre antigas práticas bibliográficas, sejam elas catalográficas e/ou classificatórias, nos permite alargar a compreensão sobre a constituição da OC no tempo presente.

A dimensão histórico-epistemológica da OC é tratada neste trabalho a partir do esquema postulado por Conrad Gesner, no Século XVI, em *Pandectarum sive partitionum universalium Conradi Gesneri Tigurini, medici & philosophiae professoris, libri XXI: Ad lectores. Secundus hic Bibliothecae nostrae tomus est, totius philosophiae*

& omnium bonarum artium atque studiorum locos communes & ordines universales simul & particulares complectens.

A partir da perspectiva histórica, é apresentado um quadro geral e introdutório da taxonomia gesneriana e o seu papel como um dos marcos à organização dos saberes e do conhecimento.

2 Conrad Gesner

Conrad Gesner (1516-1565), cientista, erudito e bibliógrafo suíço, publicou livros sobre temas múltiplos, como linguística, medicina, teologia, botânica, zoologia, paleontologia e mineralogia. Foi um dos mais notáveis cientistas do período moderno.

A sua atividade científico-erudita,

> [...] pode ser construída em torno de duas fases distintas: 1) a fase de formação em que estão incluídos os estudos clássicos, a aplicação na Medicina, o exercício da profissão, e o interesse pela Botânica [...] 2) a fase madura, ou seja, da produção científica de: a) obras bibliográficas b) obras linguístico-filológicas c) obras médicas, físicas e de ciências naturais (SABBA, 2012, p. 30).

A vocação de Gesner para elaboração de obras bibliográficas, como é o caso de *Catalogus Plantarum* (1542), resultou no projeto bibliográfico mais ambicioso da Europa Moderna: *Bibliotheca Universalis*.

A obra foi publicada em quatro partes, entre 1545-1555: 1) *Bibliotheca Universalis, sive Catalogus omnium scriptorum locupletissimus, in tribus linguis, Latina, Graeca, et Hebraica [...]* (1545); 2) *Pandectarum sive partitionum universalium [...]* (1548); 3) *Partitiones theologicae* (1549) e 4) *Appendix bibliothecae* (1555).

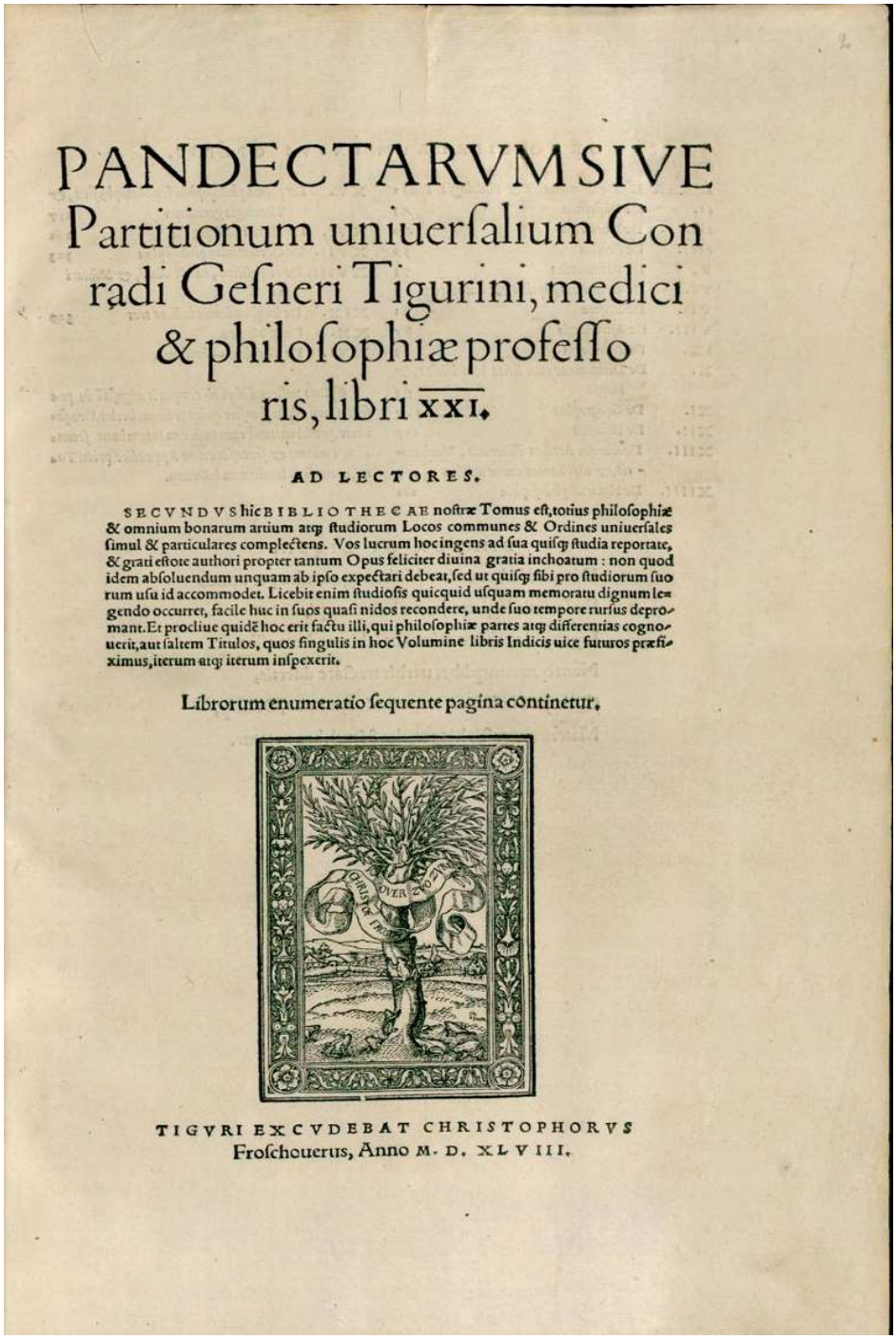
De acordo com Malclès (1960), *Bibliotheca Universalis, sive Catalogus [...]* foi impresso em 1545 por Froschover, em Zurique, em um *in-folio* de 631 folhas.

Bibliotheca Universalis, sive Catalogus [...] é um catálogo alfabético-nominal que apresenta 5031 autores de obras compostas nas três línguas eruditas, ou seja, latim, grego e hebraico. O catálogo é organizado por ordem alfabética do primeiro nome do autor, mas é acompanhado por uma lista separada e organizada por sobrenomes (SABBA, 2012). Também apresenta resumo e extratos dos documentos listados.

3 *Pandectarum Sive Partitionum Universalium [...]*

Pandectarum sive partitionum universalium [...], também conhecido por *Pandectae* (FIG. 1), foi impresso em 1548, em um *in-folio* de 374 folhas (MALCLÈS, 1960). Contempla a exposição classificada do material documentário, ou seja, a exposição do conhecimento extraído dos livros da primeira parte.

FIGURA 1 - *Pandectae*



PANDECTARVM SIVE
Partitionum uniuerſalium Con
radi Geſneri Tigurini, medici
& philoſophiæ profeſſo
ris, libri XXI.

AD LECTORES.

SECVNDVS hic BIBLIOTHECAE noſtræ Tomus eſt, totius philoſophiæ & omnium bonarum artium atq; ſtudiorum Locos communes & Ordines uniuerſales ſimul & particulares complectens. Vos lucrum hoc ingens ad ſua quiſq; ſtudia reportate, & grati eſtote authori propter tantum Opus feliciter diuina gratia inchoatum: non quod idem abſoluendum unquam ab ipſo expectari debeat, ſed ut quiſq; ſibi pro ſtudiorum ſuorum uſu id accommodet. Licebit enim ſtudioſis quicquid uſquam memoratu dignum legendo occurret, facile huc in ſuos quaſi nidos recondere, unde ſuo tempore rurſus depromant. Et procliue quidē hoc erit factu illi, qui philoſophiæ partes atq; differentias cognouerit, aut ſaltem Titulos, quos ſingulis in hoc Volumine libris Indicis uice futuros præfiximus, iterum atq; iterum inſpexerit.

Librorum enumeratio ſequente pagina continetur.

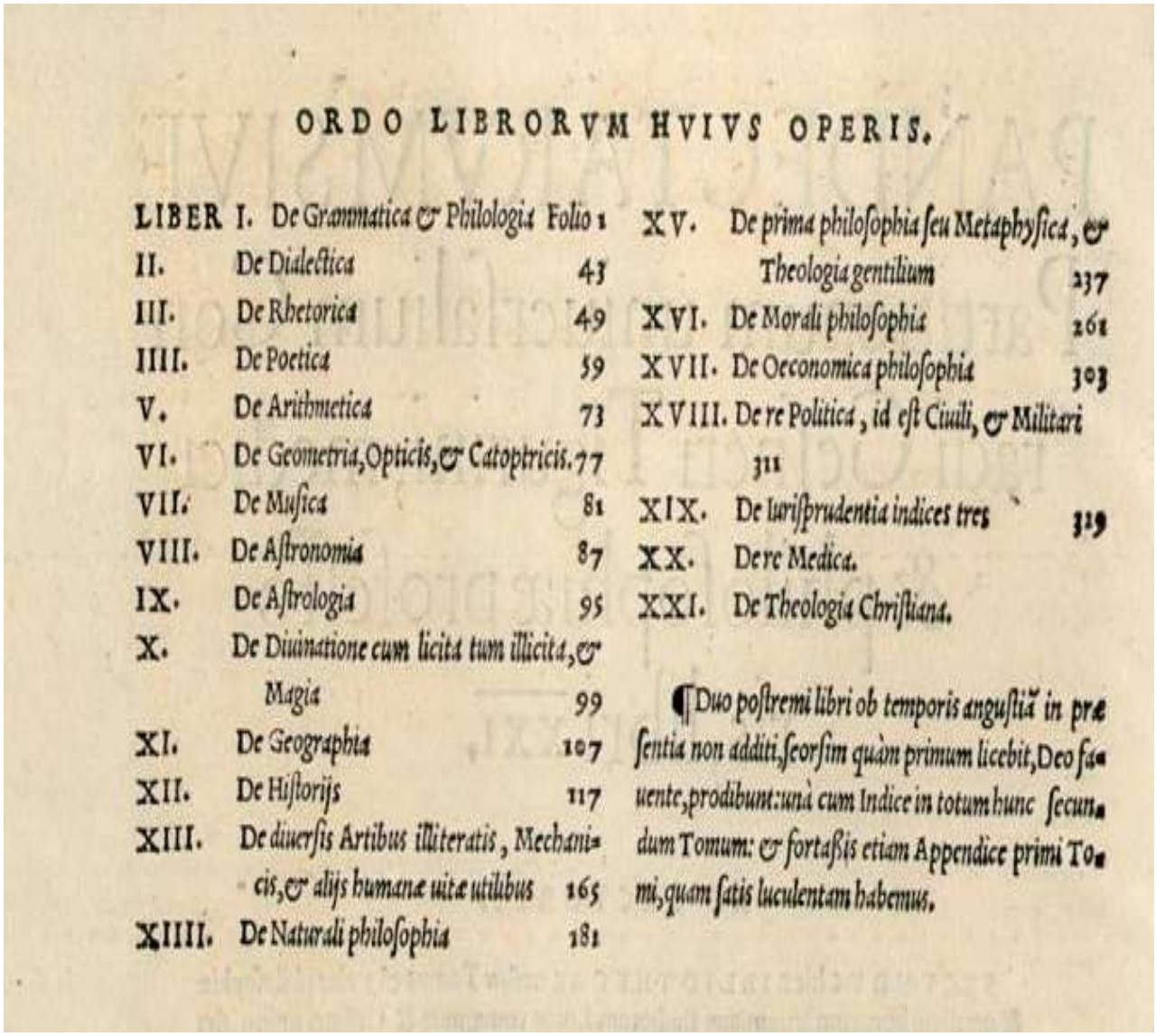
TIGVRI EXCVDEBAT CHRISTOPHORVS
Froſchouerus, Anno M. D. XLVIII.

Fonte: GESNER (1548).
Disponível em: <http://www.e-rara.ch/zuz/content/pageview/624960>

Inicialmente, o esquema foi concebido em XXI Partições: *Pandectae* contempla XIX Partições, de modo que *Partitiones theologicae* contempla a última Partição, publicada separadamente em 1549 e nomeada Partição XXI. A Partição XX, projetada para apresentar o catálogo semântico das obras de medicina, não foi finalizada (SERRAI, 1990).

O verso do frontispício de *Pandectae* apresenta o esquema completo das classes gerais da obra (FIG. 2), que resulta em um verdadeiro repertório de ordem sistemática.

FIGURA 2 – Esquema completo das classes gerais de *Pandectae*



ORDO LIBRORVM HVIVS OPERIS.

LIBER I. De Grammatica & Philologia Folio 1
II. De Dialectica 43
III. De Rhetorica 49
IIII. De Poetica 59
V. De Arithmetica 73
VI. De Geometria, Opticis, & Catoptricis. 77
VII. De Muſica 81
VIII. De Aſtronomia 87
IX. De Aſtrologia 95
X. De Diuinatione cum licita tum illicita, & Magia 99
XI. De Geographia 107
XII. De Hiſtorijs 117
XIII. De diuerſis Artibus illiteratis, Mechanicis, & alijs humanæ uitæ utilibus 165
XIIII. De Naturali philoſophia 181
XV. De prima philoſophia ſeu Metaphyſica, & Theologia gentilium 237
XVI. De Morali philoſophia 261
XVII. De Oeconomica philoſophia 303
XVIII. De re Politica, id eſt Ciuili, & Militari 311
XIX. De Iuriſprudentia indices tres 319
XX. De re Medica.
XXI. De Theologia Chriſtiana.

¶ Duo poſtremi libri ob temporis anguſtiã in præſentia non additi, ſeorſim quàm primum licebit, Deo faciente, prodibunt: unà cum Indice in totum hunc ſecundum Tomum: & fortaſis etiam Appendice primi Tomi, quam ſatis luculentam habemus.

Fonte: GESNER (1548).

Disponível em: <http://www.e-rara.ch/zuz/content/pageview/624961>

O sistema de classificação em *Pandectae* advém das sete artes liberais até as categorias de assuntos complementares e de interesse aos estudiosos e eruditos do Renascimento.

O esquema é assim organizado: 1-Gramática (e Filologia), 2-Dialética, 3-Retórica (representando o *trivium*), 4-Poética; 5-Aritmética, 6-Geometria, 7-Música e 8-Astronomia (sendo as últimas quatro classes representando o *quadrivium)*. Estas são seguidos pelas ciências incluindo o currículo universitário medieval: 9-Astrologia, 10-Adivinhação e Magia, 11-Geografia, 12-História, 13-Artes Mecânicas, 14-Filosofia Natural, 15-Metafísica, 16-Filosofia Moral, 17-Filosofia Econômica, 18-Política; e finalmente, 19-Direito, 20-Medicina e 21-Teologia.

Cada classe de *Pandectae* representa um livro que corresponde a uma partição. Cada livro, por sua vez, está organizado da seguinte forma: título do livro (associado ao nome da respectiva partição), a quem o libro é dedicado e a exposição geral das subclasses subordinadas à classe que nomeia o livro ou partição. Por fim, elenca os títulos correspondentes a cada uma das subclasses.

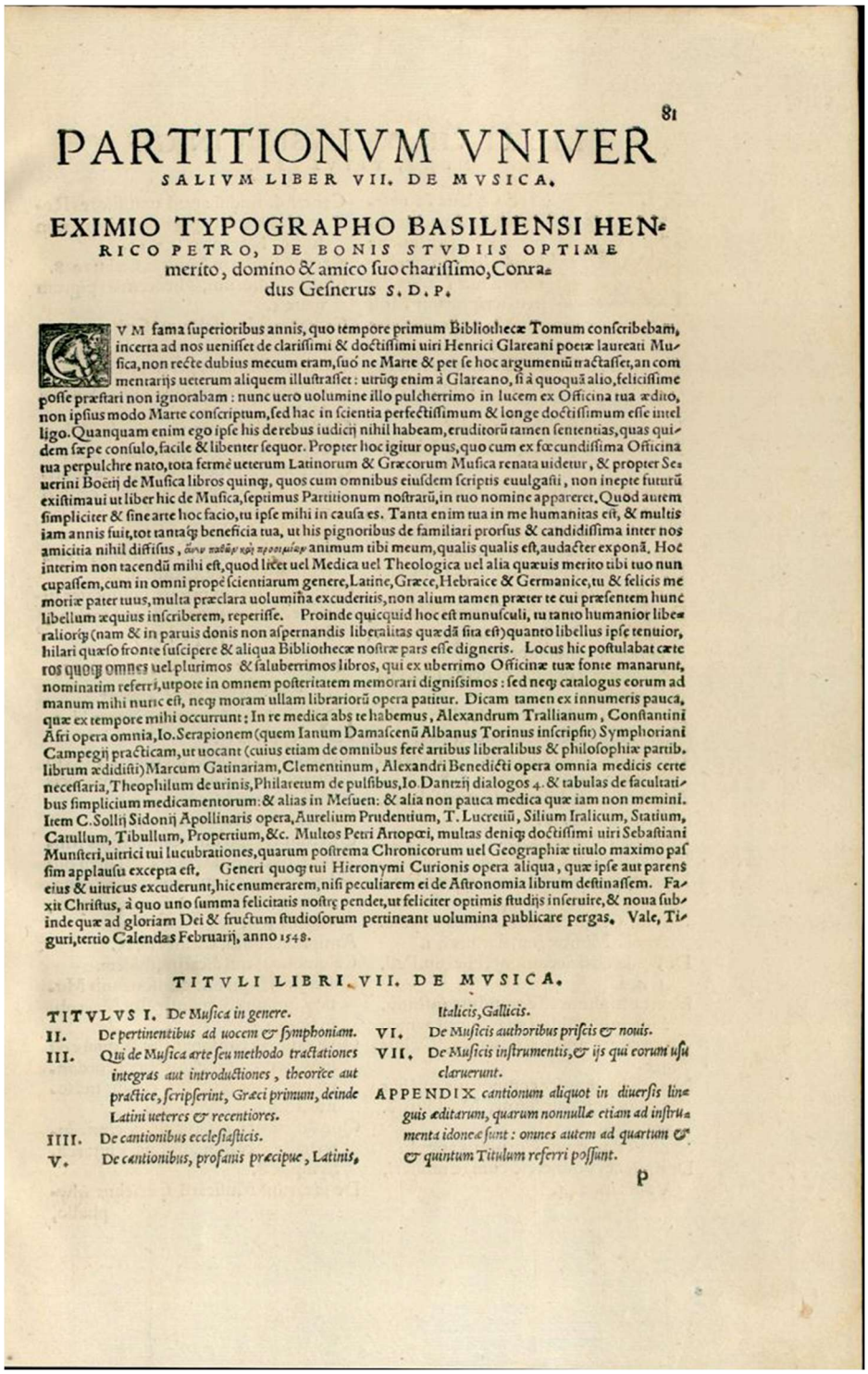
Vejamos um exemplo deste sistema a partir da Partição *De musica,* correspondente ao Livro VII*.* No anverso correspondente à página 81 de *Pandectae* (FIG. 3), lê-se as seguintes informações na sequencia:

- o título do livro/partição no topo da página: "*Partitionum universalum Liber VII. De Musica";*

- a quem o libro é dedicado: "EXIMIO TYPOGRAPHO BASILIENSI HENRICO PETRO, DE BONIS STVDIIS OPTIME merito, domino & amico suo charissimo, Conradus Gesnerus S.D.P [...]". *De musica* é dedicado a Henrich Petri ou Henricus Petrus (1508-1579), um dos maiores tipógrafos da Basiléia que encabeçou a *Officina Henricpetrina*.

- exposição geral das subclasses ou seções que constituem o Livro VII: I. Trabalhos que lidam com vários gêneros de música; II Escritos sobre teoria modal e consonância; III Tratados e materiais introdutórios que lidam com vários aspectos teóricos e práticos da música - Grego, Latim e contemporâneo; IV Edições de música litúrgica; V Edições da música secular, principalmente latim, italiano e francês; VI Autoridades antigas e recentes sobre música; VII Trabalhos sobre instrumentos musicais; VIII Livros adicionais de polifonia impressa (apêndice) (BERNSTEIN, 1973).

FIGURA 3 – *Partitionum universalum liber VII. De Musica.*



81

PARTITIONVM VNIVER

SALIVM LIBER VII. DE MVSICA.

EXIMIO TYPOGRAPHO BASILIENSI HEN-

RICO PETRO, DE BONIS STVDIIS OPTIME

merito, domino & amico suo charissimo, Conra-

dus Gesnerus S. D. P.

VM fama superioribus annis, quo tempore primum Bibliothecæ Tomum conscribebam, incerta ad nos uenisset de clarissimi & doctissimi uiri Henrici Glareani poetæ laureati Musica, non recte dubius mecum eram, suo ne Marte & per se hoc argumentū tractasset, an commentarijs ueterum aliquem illustrasset: utrūq; enim à Glareano, si à quoquā alio, felicissime posse præstari non ignorabam: nunc uero uolumine illo pulcherrimo in lucem ex Officina tua ædito, non ipsius modo Marte conscriptum, sed hac in scientia perfectissimum & longe doctissimum esse intelligo. Quanquam enim ego ipse his de rebus iudicij nihil habeam, eruditorū tamen sententias, quas quidem sæpe consulo, facile & libenter sequor. Propter hoc igitur opus, quo cum ex fœcundissima Officina tua perpulchre nato, tota fermè ueterum Latinorum & Græcorum Musica renata uidetur, & propter Seuerini Boetij de Musica libros quinq; quos cum omnibus eiusdem scriptis euulgasti, non inepte futurū existimaui ut liber hic de Musica, septimus Partitionum nostrarū, in tuo nomine appareret. Quod autem simpliciter & sine arte hoc facio, tu ipse mihi in causa es. Tanta enim tua in me humanitas est, & multis iam annis fuit, tot tantaq; beneficia tua, ut his pignoribus de familiari prorsus & candidissima inter nos amicitia nihil diffisus, ἄνευ παῤῥησίας προσιμίαν animum tibi meum, qualis qualis est, audacter exponā. Hoc interim non tacendū mihi est, quod licet uel Medica uel Theologica uel alia quæuis merito tibi tuo nunc cupassem, cum in omni propè scientiarum genere, Latine, Græce, Hebraice & Germanice, tu & felicis memoriæ pater tuus, multa præclara uolumina excuderitis, non alium tamen præter te cui præsentem hunc libellum æquius inscriberem, reperisse. Proinde quicquid hoc est munusculi, tu tanto humanior liberaliorq; (nam & in paruis donis non aspernandis liberalitas quædā sita est) quanto libellus ipse tenuior, hilari quæso fronte suscipere & aliqua Bibliothecæ nostræ pars esse digneris. Locus hic postulabat certos quoq; omnes uel plurimos & saluberrimos libros, qui ex uberrimo Officinæ tuæ fonte manarunt, nominatim referri, utpote in omnem posteritatem memorari dignissimos: sed neq; catalogus eorum ad manum mihi nunc est, neq; moram ullam librariorū opera patitur. Dicam tamen ex innumeris pauca, quæ ex tempore mihi occurrunt: In re medica abs te habemus, Alexandrum Trallianum, Constantini Afri opera omnia, Io. Serapionem (quem Ianum Damascenū Albanus Torinus inscripsit) Symphoriani Campegij practicam, ut uocant (cuius etiam de omnibus ferè artibus liberalibus & philosophiæ partib. librum ædidisti) Marcum Gatinariam, Clementinum, Alexandri Benedicti opera omnia medicis certe necessaria, Theophilum de urinis, Philaretum de pulsibus, Io. Dantzij dialogos 4. & tabulas de facultatibus simplicium medicamentorum: & alias in Mesuen: & alia non pauca medica quæ iam non memini. Item C. Solij Sidonij Apollinaris opera, Aurelium Prudentium, T. Lucretiū, Silium Italicum, Statium, Catullum, Tibullum, Propertium, &c. Multos Petri Artopœi, multas deniq; doctissimi uiri Sebastiani Munsteri, uitrici tui lucubrationes, quarum postrema Chronicorum uel Geographiæ titulo maximo passim applausu excepta est. Generi quoq; tui Hieronymi Curionis opera aliqua, quæ ipse aut parens eius & uitricus excuderunt, hic enumerarem, nisi peculiarem ei de Astronomia librum destinassem. Faxit Christus, à quo uno summa felicitatis nostrę pendet, ut feliciter optimis studijs inseruire, & noua subinde quæ ad gloriam Dei & fructum studiosorum pertineant uolumina publicare pergas. Vale, Tiguri, tertio Calendas Februarij, anno 1548.

TITVLI LIBRI VII. DE MVSICA.

TITVLVS I. De *Musica in genere.*

II. De *pertinentibus ad uocem & symphoniam.*

III. Qui de *Musica arte seu methodo tractationes integras aut introductiones, theorice aut practice, scripserint, Græci primum, deinde Latini ueteres & recentiores.*

IIII. De *cantionibus ecclesiasticis.*

V. De *cantionibus, profanis præcipue, Latinis, Italicis, Gallicis.*

VI. De *Musicis authoribus priscis & nouis.*

VII. De *Musicis instrumentis, & ijs qui eorum usu claruerunt.*

APPENDIX *cantionum aliquot in diuersis linæ guis æditarum, quarum nonnullæ etiam ad instrumenta idoneæ sunt: omnes autem ad quartum & & quintum Titulum referri possunt.*

p

Fonte: GESNER (1548).
Disponível em: <http://www.e-rara.ch/zuz/content/pageview/625132>

Na página seguinte à primeira página da Partição *De musica* tem início o elenco de títulos no escopo de cada uma das VIII subclasses desta partição.

Os livros elencados em *Bibliotheca Universalis, sive Catalogus [...]* estão dispostos em ordem alfabética de seus autores. Já em *Pandectae*, os livros elencados estão ordenados segundo os *loci communes* em razão de seus argumentos, agrupando-se a partir de seus assuntos (SERRAI, 1990).

O esquema de Gesner reflete sua concepção sobre a divisão do conhecimento diretamente atrelada ao pensamento clássico do homem renascentista.

Segundo Malclès (1960) o sistema de classificação de Gesner é original, uma vez que amplia as sete artes liberais da Idade Média.

Ao considerar os detalhes bibliográficos que passaram a constituir os livros no Séc. XVI, em função da tecnologia dos tipos móveis, *Pandectae* torna-se o primeiro esquema de classificação expressamente projetado para organizar livros.

Gesner seguiu uma lógica de revisitação às fontes uma vez que se apoiou em Johannes Trithemius (1462-1516), em sua obra *Liber de scriptoribus ecclesiasticis* (1494) - considerada o primeiro repertório biobibliográfico da Idade Moderna.

Por outro lado, Gesner inspirou o trabalho de bibliógrafos posteriores, como é o caso de Florian Treffler, O.S.B, (1483-1565) - bibliotecário e farmacêutico do mosteiro beneditino de Benediktbeuren e autor do “manual biblioteconômico” *Methodus Exhibens Variorum Indices* (1560).

Gesner organiza cada obra a partir de categorias previamente definidas e coloca em prática seu pensamento esquemático: na medida em que foi naturalista, ao trabalhar com a classificação dos seres, foi também bibliógrafo, ao trabalhar com a classificação dos saberes.

Em sua atividade científico-erudita, estava o interesse em classificar tanto livros quantos os animais. Assim, promoveu uma espécie de “anatomização” do conhecimento, sugerindo que o próprio ato de anatomizar dava bases para o ato de classificar.

4 Breve apontamento sobre possíveis contribuições de Conrad Gesner à OC

A obra de Conrad Gesner delineou as práticas bibliográficas da Europa Moderna, já que inúmeras bibliografias se basearam em *Bibliotheca.* Neste sentido, Gesner é considerado o “pai da Bibliografia” e fundador da disciplina bibliográfica.

Na contemporaneidade, Conrad Gesner é mencionado de forma breve em trabalhos sobre o tratamento descritivo e temático da informação, a exemplo de Piedade (1983), no campo da classificação; Campello (2006), no campo do controle bibliográfico; Cunha e Lima (2008), no campo do tratamento da informação jurídica; Mey e Silveira (2009), no campo da catalogação e Espanha Gomes (2010), no campo do tratamento semântico dos documentos.

Este quadro pode ser ampliado uma vez que o pensamento organizacional em Gesner é objeto inédito de investigação no contexto da OC, no Brasil, ainda que se reconheça o seu papel na fundamentação do campo.

Como nos lembra Alfredo Serrai (1990, p. 139):

> O esquema de classificação gesneriano não é simplesmente o mapa de uma série de desvios categoriais de natureza genericamente bibliográfico-literário ou bibliográfico-disciplinar; é a fachada de um edifício conceitual, enormemente complexo, cujos espaços internos foram modelados a partir de conceitos e coisas ordenadas de acordo com os modos infalíveis de uma ciência demonstrável, apoiada pela razão e sustentada pela fé.

Pandectae é fonte e objeto que constrói e evidencia, a um só tempo, um fio epistemológico que perpassa pela constituição da OC como campo, uma vez que na sua base estão os problemas de classificação que hoje são revestidos de uma nova roupagem.

Nos parece justamente ser este um dos elementos centrais da obra e que dialoga diretamente com a OC: uma reflexão sobre todas as atividades mediadoras entre a produção e o uso do conhecimento.

5 Considerações

Conrad Gesner, ao construir seu esquema, esteve inserido em um processo que marcou a Europa Moderna: a expansão da figura dos humanistas que, interessados pelos textos antigos e pelos modos em que os mesmos deveriam ser organizados, se dedicaram não só à classificação dos seres, mas também dos saberes.

O que está na base de *Pandectae* é uma postura de dissecação dos elementos das ciências naturais que se transferiu para a dissecação do conhecimento.

Como vimos, a taxonomia gesneriana se mostra inovadora em tempos de formação da Europa Moderna uma vez que amplia as sete artes liberais praticadas no Medievo, tornando-se um esquema classificatório referencial por ser o primeiro expressamente projetado para organizar livros.

O desafio que se coloca agora é voltar às ideias originais de Gesner de forma mais aprofundada, para que possamos identificar as suas implicações para OC, do ponto de vista histórico-epistemológico.

A pesquisa aponta a importância de se resgatar a figura de Gesner não só para maior compreensão dos fundamentos da OC, mas também para o necessário alargamento de estudos e debates históricos neste campo.

Referências

BERNSTEIN, Lawrence F. The Bibliography of Music in Conrad Gesner's Pandectae (1548). **Acta Musicologica**, v. 45, fasc. 1, jan./jun.1973. pp. 119-163. Disponível em: <http://www.jstor.org/stable/932225>. Acesso em: 15 maio 2015.

CAMPELLO, Bernadete. **Introdução ao controle bibliográfico**. Brasília: Briquet de Lemos, 2006.

CUNHA, Murilo Bastos da; LIMA, João Alberto de Oliveira. Tratamento da informação legislativa e jurídica: perspectiva histórica. **Senatus**: Cadernos da Secretaria de Informação e Documentação, Brasília, v. 6, n. 2, p. 33-38, out. 2008.

GESNER, CONRAD. **Pandectarum sive partitionum universalium Conradi Gesneri Tigurini, medici & philosophiae professoris, libri XXI** : Ad lectores. Secundus hic Bibliothecae nostrae tomus est, totius philosophiae & omnium bonarum artium atque studiorum locos communes & ordines universales simul & particulares complectens […].Tiguri : excudebat Christophorus Froschouerus, anno 1548. Disponível em: < http://www.e-rara.ch/zuz/content/titleinfo/624958>. Acesso em: 20 fev. 2015.

GOMES, Hagar Espanha. Longa caminhada no tratamento semântico de documentos. In: FREITAS, Lídia Silva de; MARCONDES, Carlos Henrique; RODRIGUES, Ana Célia (Org.). **Documento**: gênese e contexto de uso. Niterói: EDUFF, 2010. p. 193-204.

MALCLÈS, Louise Nöelle. **La bibliografía**. Buenos Aires: EUDEBA, 1960.

MEY, Eliane Serrão Alves; SILVEIRA, Naira Christofoletti. **Catalogação no plural**. Brasília: Briquet de Lemos, 2009.

PIEDADE, Maria Antonieta Requião. **Introdução à teoria da classificação**. 2.ed. Rio de Janeiro: Interciência, 1983.

SABBA, Fiametta. **La 'Bibliotheca Universalis' di Conrad Gesner:** monumento della cultura europea. Roma: Bulzoni Editore, 2012.

SERRAI, Alfredo. **Conrad Gesner**. Editado por Maria Cochetti. Roma: Bulzoni Editore, 1990.

SMIRAGLIA, Richard. The epistemological dimension of knowledge organization. In: DOBEDEI, Vera Dobedei; GUIMARÃES, José Augusto Chaves (Org.). **Complexidade e organização do conhecimento**: desafios de nosso século. [recurso eletrônico]. Rio de Janeiro : ISKO-Brasil ; Marília : FUNDEPE, 2013.p. 17-25. Disponível em: <http://isko-brasil.org.br/wp-content/uploads/2013/02/Estudos-avançados-2.pdf> Acesso em: 20 fev. 2015.

Configurações terminológicas e interdisciplinares na pesquisa em organização do conhecimento (2010-2014)

Leilah Santiago Bufrem
Universidade Estadual Paulista
santiagobufrem@gmail.com

Aline Elis Arboit
Universidade Estadual Paulista
aarboit@yahoo.com.br

Juliana Lazzarotto Freitas
Universidade Estadual Paulista
julilazzarotto@gmail.cm

1 Introdução

Estudos de Bufrem, Freitas e Nascimento (2014) contribuem para o delineamento de um panorama das relações entre autores, temáticas e fundamentos teóricos da área de Organização do Conhecimento (OC) por meio da análise de citação e de cocitação de artigos científicos no âmbito nacional e da identificação de temáticas e enfoques de maior representatividade. Desse modo, foram visualizadas características autorais estruturantes que compõem o cenário das pesquisas científicas no Brasil.

A possibilidade de reconhecer esse cenário foi fundamentada na convicção de que a institucionalização de qualquer área do conhecimento ocorre mediante sua consolidação teórico-empírica promovida pela comunidade científica. Assim, a busca de tendências próprias do campo em questão motivou este estudo, com o qual se

pretende ir além das considerações anteriores, aprofundando conhecimentos sobre o campo e suas relações interdisciplinares e ideológicas por meio de termos selecionados pelos autores como representativos das suas próprias obras, no que concerne aos “sentidos axiológicos" que estes termos carregam, conforme pontua Faraco (2003).

A expressiva representatividade de autores brasileiros identificados em estudo anterior reforçou a curiosidade sobre sua produção e as relações perceptíveis a partir dos termos empregados para representar essa produção. Para tanto, pretende-se analisar esses termos partindo do pressuposto de que há posições e visões de mundo distintas desses autores, embutidas nas suas preferências terminológicas e de que estas posições axiológicas podem variar se analisadas diacronicamente. A preferência por um termo em detrimento de outro possibilita a visualização de relações interdisciplinares, como efeito de opções ideológicas na OC.

2 Trajetória metodológica

A primeira etapa deste estudo consistiu de um levantamento e análise da produção científica dos autores mais profícuos em OC, a partir dos artigos de periódicos nacionais indexados na Base Brasileira de Artigos de Periódicos em Ciência da Informação (Brapci), entre os anos de 2003 e 2012.

Na segunda etapa, foram analisados os termos utilizados como palavras-chave pelos autores do corpus para representar, seus artigos publicados entre os anos de 2010 e 2014. Como fonte de dados utilizou-se a Plataforma Lattes, da qual foram coletados, por meio do Script Lattes, os artigos desses autores no período. As palavras-chave de cada autor foram identificadas e organizadas no Quadro 1 por ano e por categorias temáticas representadas por cores, as quais foram definidas simultaneamente ao percurso analítico, apresentando-se assim, as aproximações temáticas dos autores a cinco grandes temas ligados a outros campos do conhecimento, embora também tenha sido identificada pela cor azul a incidência de termos relacionados ao campo de OC. A cor amarela identifica as palavras-chave dos artigos que remetem à Epistemologia, à Metodologia da pesquisa, à Ética e aos demais temas de cunho filosófico; a verde identifica os termos relativos a Estudos métricos e análises de produção científica; a vermelha denota termos que remetem a aproximações com a Educação; a rosa identifica os termos concernentes às

Tecnologias da Informação; a marrom é utilizada para destacar termos sobre Terminologia, Linguística, Semiótica e demais Estudos de Linguagem.

Para relacionar as temáticas identificadas e a trajetória de publicação destes autores nos últimos cinco anos, definiram-se os seguintes parâmetros: a distribuição diacrônica dos termos e as suas relações paradigmáticas.

A distribuição diacrônica revela a regularidade dos temas, pois permite que se visualizem tendências temáticas no sentido de analisá-las para a recuperação do que se considera uma contribuição significativa ao tema no Brasil. Assim, trabalha-se com a sucessão dos termos, respeitando-se cronologias e processos que definem essa dimensão.

Quanto à relação paradigmática, revela a coerência entre os termos utilizados no mesmo ano, assim como entre os demais termos do período para cada autor. Concebe-se que a relação paradigmática é estabelecida pelos conceitos de um campo, que pode expressar conteúdos, métodos ou valores por eles representados. A linguística saussuriana define as relações associativas como paradigma do conjunto de elementos similares que se associam na memória e que assim formam conjuntos (SAUSSURE, 1995).

Além disso, esses termos (ou conjunto de termos) podem ser considerados signos representativos de escolhas e adesões feitas por sujeitos a determinados valores em contexto específico. Adotando-se uma concepção bakhtiniana, considera-se que esses termos, dada sua natureza, são carregados de significações ideológicas, porque tudo que é sígnico é ideológico (BAKHTIN/VOLOSHINOV, 2010). Tal entendimento justifica o estudo de tendências ou das escolhas realizadas no campo de OC que, por sua vez, estão justapostas aos termos.

3 Análise e resultados

Dos onze autores considerados mais representativos, somente sete apresentam estudos de autoria individual (26,7%). Também constatou-se, em estudo anterior de Bufrem, Freitas e Nascimento (2014), que entre esses autores, seis (54,55%) estão na liderança de grupos de pesquisa, enquanto os demais estão cadastrados como membros desses grupos. Esse resultado por si só não reflete instâncias de consagração, mas quando se considera o fato de que oito deles

(72,72%) são bolsistas de produtividade do CNPq, pode-se inferir que eles compõem a elite produtiva na área de OC no Brasil (BUFREM; FREITAS; NASCIMENTO, 2014).

O universo de autores identificados na fase exploratória foi reduzido para dez, uma vez que um deles não manteve produção no período entre 2010 e 2014. Dos demais, 50% apresentam produção anual contínua, sendo cada um de uma universidade distinta, com predomínio daquelas nas regiões Sudeste e Sul.

Bufrem, Freitas e Nascimento (2014) destacam os anos de 2010 e 2011 como aqueles de maior produção. Logo, sugere-se aqui, que o ápice da produção brasileira sobre OC nestes anos pode ser resultante do primeiro ISKO Brasil, realizado no ano de 2010.

Por meio da análise dos termos sistematizados no Quadro 1, verificaram-se na produção desses autores, além da predominância de termos ligados à OC (cor azul), outros relacionados à Tecnologia (cor rosa), como em A1 onde essa temática é preponderante, além de um elevado índice de ocorrência desse tema em A4, A8, A10 e A11. Já em A2 e A9, observa-se uma significativa interlocução com os estudos métricos (verde) e estudos de cunhos epistemológico e ético (amarelo). Maior representatividade de termos voltados para questões de linguagem, linguística, terminologia e semiótica (marrom) foi detectada em A1, A2, A4, A6 e A7. Quanto a questões educacionais (vermelho), percebeu-se uma ocorrência mais expressiva em A2.

Quando os termos categorizados são relacionados com os nomes dos grupos de pesquisa desses autores, observa-se sua convergência não somente com as palavras-chave escolhidas, mas também com relações interdisciplinares predominantes. Isso leva a crer que os artigos produzidos são também resultados de pesquisas, leituras e discussões desenvolvidas dentro dos grupos de pesquisa indicados e, desse modo, fruto de processos dialógicos e sociais contextualmente localizados.

Autores	Grupos de Pesquisa	2010	2011	2012	2013	2014
A1	* Informação e Conhecimento no Ciberespaço	Mecanismos de busca Semiótica	Memória Mecanismos de busca Ciberespaço	Mecanismo de busca Ciberespaço Tecnologias da Informação	Semiótica Mecanismos de busca Tradução intersemiótica Tecnologias da informação	Arquitetura da informação Administração pública Dobra semiótica Gilles Deleuze Web Invisível Ciberespaço
A2	* Cultura, Práticas Escolares e Educação Histórica * Educação, Pesquisa e Produção Científica	Colaboração científica Análise de citações Epistemologia Análise de citações Bases de dados Compartilhamento da informação Arquitetura da informação Brapci Periódico científico Comunicação científica Educação Biografia institucional Coautoria Recuperação da informação Método de Recuperação de Informação	Cientometria Eventos científicos Institucionalização científica Cursos de Graduação Produção científica Metodologia científica Círculo de Bakhtin Dialogismo Linguagem Lógica Currículo Ensino Conteúdo disciplinar Prática docente Didática Mikhail Bakhtin Gaston Bachelard Conceito Teoria do Conceito Terminologia Modelo Conceitual Redes sociais Bibliometria	Metodologia científica Educação infantil Cultura escolar Desenvolvimento da escrita Organização do Conhecimento Aproximações teóricas Produção periódica científica Gênero Periódico científico Bibliometria Educação do campo História Passado Significância histórica Professor Ensino médio	Produção científica Opções metodológicas Brapci Pesquisa científica Métodos de pesquisa Editoras universitárias Práticas editoriais Educação do campo Biblioteca Escolar Leitura Escrita Institucionalização da Ciência da Informação Bibliometria Estudos de Citação GT 7 do ENANCIB Brasil Escrita Literária Letramento Literário Ensino Fundamental	Organização do conhecimento Pesquisa científica Literatura periódica científica Pesquisa em organização do conhecimento Autoria em organização do conhecimento Análise de citação Análise de cocitação Produção científica Editoras universitárias Práticas editoriais Editores universitários Ciência da Informação Análise da produção científica Acoplamento bibliográfico Metodologia da pesquisa Educação do campo Biblioteca Escolar Leitura Escrita Paraná Curso de Pedagogia Disciplina de PPP Disciplinas Escolares Código Disciplinar Práxis Pedagógica
A4	* Bioinformática e Genética Molecular Evolutiva * Ontologia e Taxonomia: Aspectos Teóricos e Metodológicos * Modelagem Conceitual para Organização Hipertextual de Documentos – mhtx.	Terminologia Abordagem Onomasiológica Abordagem Semasiológica Definição Compatibilização semântica	Extração de informação Reuso de ontologia Alinhamento de ontologia	-	Organização do conhecimento Teorias de Representação Categorização Análise de Domínio Anotação Genômica Reuso de ontologia Alinhamento de ontologia Compatibilidade de linguagem Tripanosomatídeos Aquisição de conhecimento	--
A5	* RESNAPAP * TEMMA * SCIENTIA		Institucionalização científica Cursos de Graduação Turismo Brasil		Bibliometria Cientometria Indicadores de CT&I Visualização de Informação	
A6	* TEMMA			Recuperação da informação		Pesquisa terminológica Tesauros Documentos artísticos Museus

				Bem cultural Terminologia Linguística Ferdinand de Saussure Linguística documentária		
A7	* TEMMA * Observatório da Museologia Baiana	Recuperação da informação Bases de dados Estruturação da informação Sistemas de informação		Internacionalismo Sistemas de recuperação da informação Vocabulários semânticos Web semântica Interoperabilidade semântica Categorização Categorias	Interoperabilidade semântica Web semântica Vocabulários semânticos Visibilidade Bases de dados„ Intersecção entre bases de dados	José María Izquierdo Arroyo Semiótica documental Linguística documental Lingüística documentária Semiótica documentária
A8	* Organização/Produção do Conhecimento Científico e Institucionalidades Positivas, * LES - Laboratório de Engenharia de Software * Organização do Conhecimento para Recuperação da Informação * Ontologia e Taxonomia: Aspectos Teóricos e Metodológicos * Gestão e Uso da Informação e do Conhecimento	Terminologia Abordagem onomasiológica Abordagem semasiológica			Reuso de ontologia Alinhamento de ontologia Compatibilização de linguagem Tripanosomatídeos Aquisição de conhecimento Etnometodologia Anotação genômica	
A9	* TEMMA * Formação e atuação profissional em organização da informação * Análise Documentária * Estudos métricos em informação	Copyright Exceções ao copyright Ibero-America Países em desenvolvimento Terminologia Abordagem onomasiológica Abordagem semasiológica Deontologia Ética Ética Profissional	Análise bibliométrica Indicadores de citação Redes de citação Ética Bias Não neutralidade Problemas éticos Redes Sociais Colaboração Científica Coautoria Diplomática Contemporânea	Frentes de pesquisa Brasil América do Sul Homossexualidade masculina Ética Brasil	Universalidade Pós-estruturalismo	Tecnologia da comunicação Cientistas da informação Produção científica Colaboração científica Bolsistas de Produtividade em Pesquisa do CNPq Análise de domínio Produtividade científica Redes de colaboração Brasil Estados Unidos Ibéria Indexação sistemática Método analítico-sintético
A10	* Núcleo de Estudos em Informação e Mediações Comunicacionais Contemporâneas * Comunicação e Divulgação Científicas	Web		Ontologia Recuperação da informação Web semântica		FRSAD Comércio eletrônico Navegação Web

			Teorias Semânticas Análise Conceitual	Ontologia Web semântica Aprendizado de ontologia Mineração de textos Sistema de informação		Taxonomia navegacional Relações semânticas
A11	* Núcleo de Estudos em Informação e Mediações Comunicacionais Contemporâneas * Estudos em Organização do Conhecimento: Abordagens Teóricas, Metodológicas e Historiográficas * Estudos de Representação e Organização da Informação e do Conhecimento	Repositórios Objetos de Aprendizagem Metadado Dublin Core ETD-MS MTD-BR BDTD NDLTD	Organização do conhecimento Representação do conhecimento (Sistemas especialistas) Semântica	Web Semântica Semiótica	Metadados Informação musical	Terminologia Museus Patrimônio digital Exposição virtual

Quadro 1: Termos utilizados pelos autores da OC (2010-2014).
Fonte: As autoras, 2015.

4 Considerações finais

A análise diacrônica dos termos revela domínios predominantes que interagem com a OC, indicando também que estes domínios incidem de modo equilibrado ao longo dos anos, e que a OC não prioriza um em detrimento de outro em períodos específicos. Este fato denota o constante trabalho interdisciplinar entre autores da área e a variedade de enfoques e frentes de pesquisa potenciais para o campo.

Desse modo, a validade deste estudo residiu na possibilidade de articulação da investigação empírica com o arcabouço teórico e interdisciplinar passível de ser visualizada a partir das relações estabelecidas. Entre elas, a concordância das palavras-chave utilizadas e as temáticas estudadas nos grupos de pesquisa dos autores indica uma aproximação da produção científica destes com suas linhas e grupos. Ainda é notável que as relações associativas entre os termos dos mesmos autores destacam-se como um indício de coerência desta produção científica e de aprofundamento das temáticas estudadas no decorrer dos anos. Sugere-se, portanto, que essa aproximação pode ser considerada um ponto forte na produção que representa esses autores, tendo em vista que a participação legitimada em grupos de pesquisa presume possibilidades de aprofundamento e continuidade de pesquisas em coautoria entre os membros do mesmo grupo.

Referências

BAKHTIN, M. (VOLOCHÍNOV). **Marxismo e filosofia da linguagem**. 14. ed. São Paulo: Hucitec, 2010.

BUFREM, L. S.; FREITAS, J. L.; NASCIMENTO, B. S. Autoria e pesquisa em Organização do Conhecimento: análise da produção científica em Ciência da Informação. **Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS**, Porto Alegre, v. 20, n. 3, ed. especial, 2014.

FARACO, C. A. **Linguagem e diálogo**: as ideias lingüísticas do círculo de Bakhtin. Curitiba: Criar, 2003.

FREITAS, J. L.; BUFREM, L. S.; NASCIMENTO, B. S. A organização do conhecimento na dinâmica da pesquisa em artigos da literatura científica da Brapci. **Transinformação**, Campinas, v. 26, n. 3, 2014.

SAUSSURE, F. **Curso de lingüística geral**. São Paulo: Cultrix, 1995.

A dimensão conceitual da organização do conhecimento nos congressos da NASKO: análise de conteúdo Bardiana

José Augusto Chaves Guimarães
Universidade Estadual Paulista
guima@marilia.unesp.br

André Ynada dos Santos
Universidade Estadual Paulista
andre.ynada@gmail.com

Rodrigo de Sales
Universidade Federal Fluminense
rodrigo.sales.s@gmail.com

Daniela Fernanda de Oliveira Matos
Universidade Estadual Paulista
danielamatos@marilia.unesp.br

1 Introdução

A organização do conhecimento desempenha papel nuclear na Ciência da Informação, apresentando avanço teórico notadamente a partir da criação da ISKO, em 1989. No entanto, observa-se que esse campo ainda se encontra em fase de consolidação, em busca da construção e delimitação de sua própria base conceitual e seus limites, em virtude de distintas influências teóricas. Tal aspecto se revela, dentre outros, por uma diversidade no tocante ao seu aspecto conceitual, notadamente com relação a seu objeto, o que leva à necessidade de analisar de que forma o ambiente científico da ISKO vem construindo / delimitando essa dimensão conceitual, a partir de seu discurso científico oficial em âmbito internacional.

Assim, pesquisa realizada por Guimarães et al (2014), relativa ao universo dos congressos internacionais da ISKO, demonstrou que a organização do conhecimento constitui área ou campo de conhecimento e/ou de aplicação onde, relacionado à

construção de discursos especializados. Sua natureza operacional encontra-se ora voltada à organização e representação de conceitos, ora voltada à busca de informação e ao acesso aos conhecimentos, em especial diálogo com a recuperação da informação.

Tendo por objeto a estrutura do conhecimento contida nos documentos (conhecimento registrado, socializado e publicado), com especial ênfase aos conceitos e à sua modelagem assim como às atividades discursivas em domínios científicos e práticas sociais e culturais específicos, a organização do conhecimento vale-se de instrumentos como *sistemas de classificação,* tesauros, vocabulário e outras linguagens de indexação. Nesse contexto, as abordagens predominantes relacionam-se às questões cognitivas, tecnológicas, lógico-conceituais, socioculturais e de gestão, o que sinaliza para os diálogos mais premente na área.

Nesta oportunidade, e em continuidade à referida pesquisa, analisa-se a questão conceitual da organização do conhecimento nos congressos realizados pelo capítulo norte-americano da ISKO.

A vista disso, parte-se de uma abordagem sucinta da organização do conhecimento enquanto campo de estudos para chegar ao contexto da International Society for Knowledge Organization – ISKO, seja em sua organização internacional, seja me seus capítulos nacionais ou regionais.

Em seguida, descreve-se a forma de coleta de dados, a partir dos anais dos congressos do capítulos norte-americano da ISKO, os quais são apresentados, analisados e discutidos sob duas das onze vertentes de análise de domínio propostas por Hjorland (2002): a) epistemológica, a partir da análise de conteúdo das definições apresentadas, e bibliográfica, a partir das redes de citação que subsidiam essa produção.

No âmbito da Ciência da Informação, a organização e representação do conhecimento apresenta natureza mediadora, configurando-se em um conjunto de processos que estabelecem a intermediação entre um conhecimento que, uma vez produzido, foi materializado e socializado, de tal forma que esse conhecimento possa servir de base para a geração de um novo conhecimento. Este novo conhecimento, por sua vez, uma vez materializado e socializado, pode igualmente ser objeto de nova organização e representação caracterizando aquilo que se pode denominar como fluxo helicoidal da informação (GUIMARÃES, 2008).

Desse modo, como bem recorda GARCIA MARCO (1997, p.8) a Organização do conhecimento surge como ramo de estudos mais notadamente a partir da década de 90 do século XX,

> [...] na encruzilhada das denominadas ciências cognitivas, no campo de encontro entre as Ciências do Conhecimento Humano (Neurociência, Psicologia e Epistemologia), Ciências da Informação e da Comunicação (incluindo a Semiótica e a Linguística), matemática (incluindo a Lógica e as Linguagens formais) e a Ciência da Computação.

O termo *Organização do Conhecimento* foi introduzido por Henry Evelyn Bliss, a partir de suas obras "The organization of Knowledge and systems of sciences", de 1929, e "The organization of knowledge in libraries", de 1933. Referindo-se à precursora obra de Evelyn Bliss. *A organização do conhecimento e o sistema das ciências*, de 1931, Ingetraut Dahlberg (1993, p.211) afirma que:

> Lidamos com um volume de conhecimento que foi coletado, aprofundado e amadurecido ao longo de séculos o que, no entanto, apenas em nossos dias tem sido reconhecido como uma área autônoma de conhecimento necessitando encontrar seu próprio lugar na sociedade de modo a ser reconhecida no âmbito do sistema das ciências.

Posteriomente, o termo foi resgatado e discutido por Dagobert Soergel (1971) e Ingetraut Dalhberg (1993).

Para Dahlberg (2008), a Organização do Conhecimento possui natureza científica e destina-se a sistematizar unidades de conhecimento (conceitos) a partir dos elementos ou características que lhes são inerentes. A isso se alia a aplicação de conceitos e de classes de conceitos que possam permitir a transferência de conteúdos (assuntos). Para a autora (DALHBERG, 1993) duas concepções de conhecimento devem ser consideradas: enquanto processo individual, que não é transferível, e caquele que pressupõe um consenso, a partir de um registro, sendo, este último, objeto da organização do conhecimento. Assim, o termo Organização do Conhecimento, para a Ciência da Informação irá se dedicar às atividades de classificar, indexar e representar o conhecimento através de registros com o objetivo específico de suprir a necessidade de informação (BUFREM, 2008).

No âmbito investigativo, a Organização do Conhecimento tem se dedicado a estudar leis, princípios e procedimentos nos quais se sustenta o conhecimento, em cujo contexto se destaca a atuação da International Society for Knowledge Organzation – ISKO, instituída em 1989, na Alemanha, por iniciativa de Ingetraut Dahlberg como uma sociedade de pesquisadores que abordam a temática de Organização do Conhecimento. Sua missão consiste em promover o avanço teórico e aplicado da Organização do Conhecimento em diversos campos e em diferentes maneiras, congregando pesquisadores de distintos países, em áreas como Ciência da Informação, Linguística, Filosofia, e Ciência da Computação entre outras.

Em termos de atividades científicas da ISKO, destacam-se as conferências internacionais, que ocorrem nos anos pares, as conferências nacionais ou regionais em anos ímpares, além de duas publicações grande impacto científico na área: a revista trimestral Knowledge Organization e a série Advances in Knowledge Organization.

Ponto alto da atuação científica da ISKO, as conferências internacionais vêm se realizando bienalmente na seguinte conformidade: Darmstadt, 1990; New Delhi, 1992: Copenhagen, 1994; Washington,1996; Lille, 1998: Toronto, 2000; Granada, 2002; Londres, 2004; Viena, 2006; Montréal, 2008; Roma, 2010; Mysore, 2012; e Cracóvia, 2014, sendo que o próximo congresso internacional, em 2016, realizar-se-á no Rio de Janeiro.

Em suas relações internacionais, a ISKO mantém atividades colaborativas com a UNESCO, a Comissão Europeia, a ISO (International Standardization Organization), a IFLA (International Federation of Library Associations), a ASIST (American Society for Information Science and Technology), entre outras.

Em termos operacionais, a ISKO desenvolve suas atividades a partir de capítulos nacionais ou regionais que, por sua vez, realizam congressos bienais. Hoje, tem-se os seguintes capítulos: Germânico, Norte-americano, Francês, Espanhol, Italiano, Brasileiro, Polonês, Escandinavo, Maghreb e Índia.

A ISKO América do Norte – NASKO, criada em 2007, é formada pelo Canadá e pelos Estados Unidos, tendo realizado até então quatro congressos sobre os temas: Knowledge Organization (Toronto, 2007), Pioneering North American Contributions to Knowledge Organization, (Syracuse, 2009); Expanding Our Horizons, Evaluating Our

Parameters (Toronto, 2011) e Cultures, Transition KO: Evolving Exploration, Critical Reflection, and Practical Work (Milwaukee, 2013).

2 Metodologia

A pesquisa desenvolveu-se a partir do conjunto de atas dos congressos do capítulo norte-americano (2007, 2009, 2011 e 2013) da ISKO (International Society for Knowledge Organization). A escolha dos referidos capítulos deu-se pelo fato de um capítulo que, embora recente, revela-se bastante produtivo na área em âmbito internacional e com forte presença nos congresoss internacionais da ISKO e na revista Knowledge Organization.

A pesquisa centrou-se na abordagem da análise de domínio que, no âmbito internacional da Ciência da Informação, vem sendo tradicionalmente trabalhada, em termos teóricos e aplicados, por Hjorland & Albrechtsen (1995); Moya Anegón & Herrero Solana (2001), Hjorland (2002, 2004), Tennis (2003), e Smiraglia (2011), dentre outros, constitui uma importante abordagem para caracterização e avaliação da ciência, na medida em que permite identificar as condições pelas quais o conhecimento científico se constrói e se socializa. Para tanto, tem-se por domínio a concepção de Smiraglia (2012, p.114) enquanto "… um grupo que apresenta uma base ontológica que revela uma teleologia subjacente, um conjunto de hipóteses comuns, assim como um consenso epistemológico sobre as abordagens metodológicas, sociais e semânticas a serem utilizadas".

Desse modo, valeu-se de duas das onze abordagens de análise de domínio previstas por Hjorland (2002): a epistemológica e a bibliométrica.

Para tanto, e tendo por base o percurso metodológico utilizado por Guimarães, Martínez Ávila, Sales e Alencar (2014) partiu-se de um levantamento na coleção completa dos referidos Anais, tomando por base a incidência do termo "knowledge organization" (e seus correspondentes em francês, espanhol e português) suas variações em francês, nos títulos, palavras-chave, resumos e títulos de seções das comunicações publicadas.

Em seguida, realizou-se a leitura das comunicações que revelaram tal incidência, selecionando-se aquelas que traziam conceitos, definições ou considerações teóricas acerca da natureza da organização do conhecimento.

A dimensão epistemológica foi abordada a partir do universo conceitual da literatura analisada. Dessa forma, no tocante às definições, procedeu-se à Análise de Conteúdo (Bardin, 2003), cuja escolha se justifica pelo fato de seus procedimentos possibilitarem uma análise com base em inferências extraídas de conteúdos de documentos – a partir de uma interpretação controlada por meio de variáveis ou indicadores, que proporcionam maior liberdade ao analista, sem que se perca a objetividade da investigação, e por se tratar de uma análise pautada em definições concebidas por autores/pesquisadores, ou seja, argumentos registrados textualmente. Dividida em três fases: pré-análise, exploração do material e, tratamento dos resultados (inferências e interpretações). Para tanto, a delimitação do corpus respeito as quatro regras básicas previstas por Bardin (2003): exaustividade (na medida em que todas as definições encontradas, quando da coleta de informações, foram consideradas materiais úteis à análise, sem processo seletivo); representatividade (assegurada pela escolha do canal de comunicação científica, pois os trabalhos publicados nos Anais da ISKO são de autoria de pesquisadores e teóricos que representam uma parcela significativa da produção internacional em organização do conhecimento); homogeneidade (pois as definições analisadas são relativas a um mesmo tema, assim, são homogêneos tanto na estrutura textual quanto no assunto); e pertinência do material (as fontes são as próprias ideias explicitadas pelos autores/pesquisadores em suas definições).

As definições foram analisadas a partir das seguintes categorias: Natureza; Objeto, Objetivo, Instrumentos, Processos, Diálogos interdisciplinares e Perspectivas ou abordagens.

Em seguida, e de modo a caracterizar o domínio, realizou-se uma análise bibliométrica, mais especificamente no que se refere à análise de citações, a partir das fontes (referentes teóricos) utilizadas pelos trabalhos selecionados, de modo a identificar dialógicos, consonâncias correntes teóricas, etc.

Por meio da análise de citações torna-se possível analisar o impacto e a visibilidade de autores em um dado domínio científico, sendo especialmente útil para que se identifiquem escolas teóricas. Dessa forma, constitui especial ferramenta para que, a partir de análise quali-quantitativas, se possa melhor compreender o universo epistemológico de uma dado domínio.

A partir desses estudos torna-se possível a construção e visualização gráfica de redes sociais de citação, especialmente útil na visualização do comportamento de um dado domínio de conhecimento.

Para fins de visualização das redes de citação, utilizou-se do software PAJEK, descrito por Adamic (2008).

3 Apresentação, análise e discussão dos resultados

Partiu-se de um conjunto de 66 trabalhos na NASKO. Desse conjunto, e a partir da incidência dos termos no título, chegou-se a um corpus de análise de 51 trabalhos (Abbas, 2007; Adler & Tennis, 2013; Beak et al, 2013; Bernstein, 2009; Buchel & Hill, 2009; Campbell, 2009, 2011; Campbell et al., 2007; Dousa, 2009a, 2009b, 2011; Edwards, 2011, 2013; Feinberg, 2007; Fox, 2011; Fox & Reece, 2013; Frické, 2013; Graf & Smiraglia, 2013; Green, 2009, 2011; Green & Fallgren, 2007; Guimarães & Sales, 2013; Heuvel, 2011; Hoffman, 2009; Hudon, 2011; Hudon & Guitard, 2013; Jacob & Zhang, 2013; Kasten, 2007; Kemp, 2007; La Barre, 2009; Loehrlein, 2011; MacCall, 2011; Marchese & Smiraglia, 2013; Martínez-Ávila & Smiraglia, 2013; McTavish & Fortier, 2011; Milani & Guimarães, 2011; Moulaison; Dykas & Budd 2013; Olson & Howarth, 2013; Pattuelli, 2011; Pimentel, 2007, 2009, 2011; Schatz; Stennes-Spidahl; Mills & Loehrlein, 2013; Smiraglia, 2007, 2009, 2013; Szostak, 2013; Tennis, 2007; Thornton, 2011; Tognoli; Guimarães & Tennis, 2013; and Zhang, Kipp & Lee, 2013).

Partindo do pressuposto que uma análise de caráter terminológico não busca unanimidades, mas sim concepções que sobressaiam diante de um corpus referente a um assunto analisado, é possível destacar, de acordo com a maioria dos textos analisados, que a organização do conhecimento vem sendo predominantemente abordada pelos pesquisadores da NASKO numa perspectiva instrumental sistemática, aja vista que a incidência da abordagem da KO como um sistema, ou uma forma sistemática de estruturas de conhecimentos, é a abordagem mais recorrente nos textos investigados. É possível observar no quadro acima que, no universo analisado, em 26 papers se destaca o fato da organização do conhecimento estar vinculada à questão dos sistemas de organização do conhecimento. Tal fato não significa necessariamente que os autores entendam a KO como um sistema somente, mas sim, que abordar os aspectos

inerentes à organização do conhecimento, na maioria das vezes, significa destacar seu caráter sistemático de estruturação do(s) conhecimento(s).

É necessário ponderar o fato de que as variáveis apresentadas acima não são excludentes, pois, ao observar atentamente, é possível verificar que muitos autores empregam diversas perspectivas ao versar sobre KO. Como é o caso do texto escrito por Olson & Howarth (2013), por exemplo. O referido texto aborda a KO sob naturezas diversas, enquanto sistema, enquanto tema especializado, enquanto processo ou, ainda, enquanto comunidade. A incidência de textos que abordam a KO sob diferentes perspectivas é mais comum que o esperado. E isso pode revelar, dentre outras coisas, pelo menos dois fatos: a natureza multifacetada da organização do conhecimento e a incipiente discussão sobre sua natureza.

O termo organização do conhecimento foi vinculado aos sistemas de organização do conhecimento nos seguintes trabalhos: Abbas, 2007; Feinberg 2007; Green & Fallgren, 2007; Pimentel, 2007; Kasten, 2007; Bernstein, 2009; Buchel &Hill, 2009; Pimentel, 2009; Green, 2009; Campbell, 2009; Green, 2011; Heuvel, 2011; Edwards, 2011; Pattuelli, 2011; Pimentel, 2011; Fox, 2011; McTavish & Fortier, 2011; Martinez-Avila & Smiraglia, 2013; Moulaison, Dykas & Budd 2013; Graf & Smiraglia, 2013; Adler & Tennis, 2013; Fox & Reece, 2013; Frické, 2013; Guimaraes & Sales, 2013; Olson & Howarth, 2013; Schatz; Stennes-Spidahl, Mills & Loehrlein, 2013.

A organização do conhecimento entendida como um processo figura nos seguintes trabalhos: Campbell et al, 2007; Kasten, 2007; Pimentel, 2007; Kemp, 2007; Tennis, 2007; Hoffman, 2009; Fox, 2011; MacCall, 2011; Milani & Guimarães, 2011; Adler & Tennis, 2013; Fox & Reece, 2013; Edwards, 2013; Jacob & Zhang, 2013; Olson & Howarth, 2013.

Porém, a organização do conhecimento já vem sendo tratada como um campo investigativo e é apresentada nos trabalhos de: Feinberg 2007; Kasten, 2007; Smiraglia, 2007; Smiraglia, 2009; La Barre, 2009; Bernstein, 2009; Campbell, 2011; Loehrlein, 2011; Milani & Guimarães, 2011; Zhang, Kipp & Lee, 2013; Adler & Tennis, 2013; Marchese & Smiraglia, 2013; Tognoli; Guimarães & Tennis, 2013.

Em seguida, verifica-se a organização do conhecimento abordada como a) uma comunidade: Pimentel, 2007; Campbell, 2009; Smiraglia, 2009; Edwards, 2013;

Moulaison, Dykas & Budd 2013; Thornton, 2013; Olson & Howarth, 2013; b) como um domínio: Smiraglia, 2009; Beak et al, 2013; Graf & Smiraglia, 2013; Smiraglia, 2013; Guimaraes & Sales, 2013 ou; c) como um tema de especialidade: Dousa, 2009; Hudon, 2011; Hudon & Guitard, 2013; Olson & Howarth, 2013.

As interpretações e inferências revelam que, embora haja uma predominância ao vincular a KO aos sistemas de organização do conhecimento, a multifacetagem é uma característica marcante na conceituação da KO no âmbito da NASKO.

Após a abordagem da KO enquanto sistema (presente em 26 textos), as perspectivas mais ocorridas foram da organização do conhecimento enquanto processo (ou conjuntos de práticas e procedimentos), presente em 14 textos e, enquanto campo de estudo, apresentado em pelo menos 13 textos analisados. Em seguida, as perspectivas mais flagrantes nos textos foram as da KO enquanto comunidade científica (em 7 textos), enquanto domínio (em 5 textos) e enquanto tema especializado (em 4 textos).

Levando a 593 autores citados (excluídas as entidades coletivas e as auto-citações), e autores que receberam duas ou mais citações: 159 (18,5%), no total de 752 citações.

Aplicando-se a Teoria do Elitismo de Price, chega-se a um grupo de 27 autores, o que recai em um autor com 5 citações. Desse modo, foram considerados todos os autores citados ao menos 5 vezes, que levou a um grupo de 29 autores.

Com base em tais aspectos, elaboram-se redes de citação por meio do software Pajek, como se pode verificar a seguir:

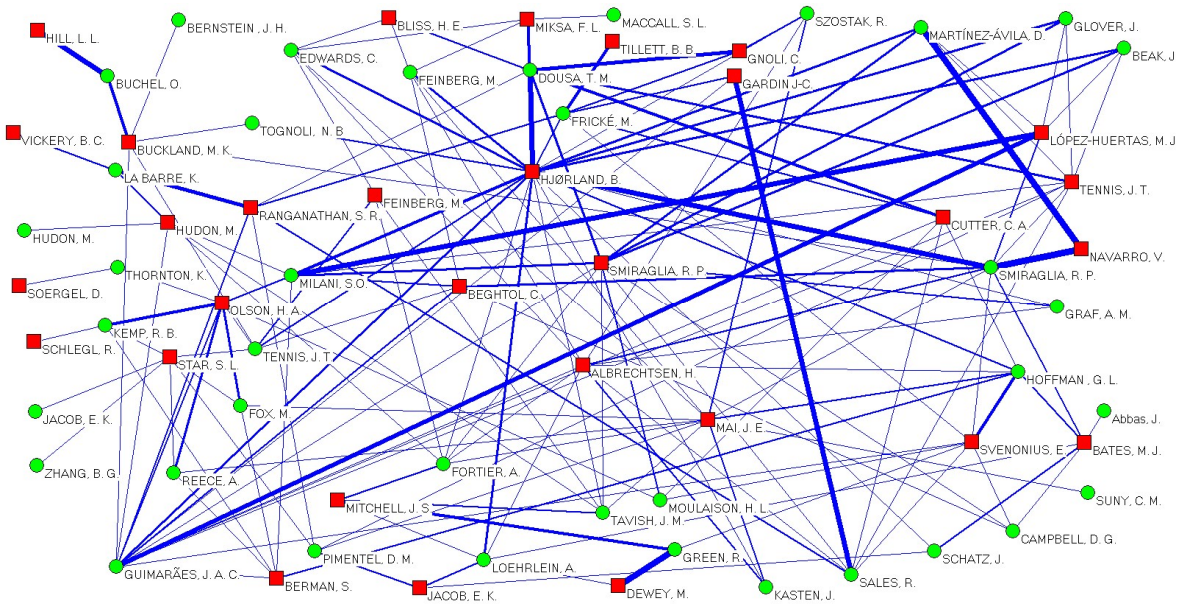


Figura 1: Rede de citação. Fonte: Elaborado pelos autores
Legenda: Círculos Verdes- Citantes; Quadrados Vermelhos- Citados.

Entre os autores citados, destacam-se o número de pesquisadores que citam B. Hjørland, R.P. Smiraglia, H.A. Olson, M.J. Lópes-Huertas e V. Navarro evidenciados pelo grande número de vetores dos citantes em relação a esses citados.

As relações mais fortes entre os investigadores citantes e autores citados são observadas de J.A.C. Guimarães, S.O. Milani, para M.J. Lópes-Huertas, R. Sales para J-C. Gardin e D. Martínez-Ávila, R.P. Smiraglia para V. Navarro, evidenciando assim o forte enlace desta comunidade citante em relação aos autores citados. Ainda, destaca-se a intensa citação de Hjørland, por J.A.C. Guimarães, S.O. Milani e R.P. Smiraglia.

4 Considerações finais

Se por um lado os resultados da análise não permitem assertivas mais definitivas, pelo fato das variáveis não serem excludentes, não tornando possível estabelecer, assim, como a comunidade de KO norte-americana concebe a natureza da organização do conhecimento, por outro, torna-se possível inferir que seja qual for a perspectiva adotada ao conceituar a KO, sua ligação com os sistemas de organização do conhecimento ou

com os esforços de sistematização do conhecimento é fortemente tratada pelos pesquisadores norte-americanos.

Tradicionalmente, a corrente norte-americana alavancou, em finais do século XIX e inícios do XX, o discurso científico da padronização de procedimentos de tratamento da informação para a construção de produtos e serviços próprios da biblioteconomia, que repercutiram fortemente no cenário internacional. Pode-se afirmar que, de certo modo, isso se reflete ainda neste novo espaço denominado KO, pois os sistemas e os processos ainda são dominantes na literatura. Porém, não é difícil perceber que aquela ótica mais pragmática já vem, há certo tempo, dividindo espaço, no âmbito da NASKO, com preocupações epistemológicas mais acentuadas, fazendo emergir perspectivas que preferem tratar a organização do conhecimento como campo, domínio ou comunidade científica.

Referências

ABBAS, J. (2007). In the margins: Reflections on scribbles, knowledge organization, and access (extended abstract). In Tennis, J. T., Ed. *Proceedings of the North American Symposium on Knowledge Organization, 1.* Toronto, pp. 107-111. Available: http://iskocus.org/nasko2007-proceedings.php.

ADAMIC, L. Introductory social network analysis with Pajek. Ann Arbor (MI): School of Informaiton. University of Michigan, 2008. Disponível em: http://ocw.mit.edu/courses/economics/14-15j-networks-fall-2009/assignments/MIT14_15JF09_pajek.pdf. Acesso em 14.07.2014.

ADLER, M.; TENNIS, J. (2013). Toward a taxonomy of harm. In: *Proceedings from North American Symposium on Knowledge Organization, v. 4*. Milwaukee: University of Wisconsin-Milwaukee, pp. 1-19.

ANDERSEN, J. (2002). Communication technologies and the concept of knowledge organization. *Knowledge Organization* 29: 29-39.

BARDIN, L. L' analyse du contenu. 7. ed. Paris: PUF, 2003. 296 p.
BEAK, J., GLOVER, J., MARTÍNEZ-ÁVILA, D., MILANI, S.O. (2013). International Comparative Domain Analysis in Knowledge Organization Research Topics in Four Countries - Brazil, South Korea, Spain and the United States. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

BEAK, J.; OLSON, H. A. (2011). Analysis of metadata schemas for children's libraries. In: Smiraglia, R. P., ed., *Proceedings from North American Symposium on Knowledge Organization,, 3*, Toronto, pp. 13-22. Available: http://iskocus.org/nasko2011-proceedings.php.

BERNSTEIN, J. H. (2009). The data-information-knowledge-wisdom hierarchy and its antithesis. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 68-75. Available: http://iskocus.org/nasko2009-proceedings.php.

BUCHEL, O.; HILL, L. (2009). Treatment of Georeferencing. in Knowledge Organization Systems: North American Contributions to Integrated Georeferencing. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 47-57. Available: http://iskocus.org/nasko2009-proceedings.php.

BUFREM, L.S. Práticas de organização e divulgação da produção intelectual em ciência da informação no Brasil. Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis: v. 13, n. esp., p. 36-53, 1º sem. 2008.

CAMPBEL, D.G. (2011). RDA and RDF: A discourse analysis of two standards of resource description. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 207-216. Available: http://iskocus.org/nasko2011-proceedings.php.

CAMPBELL, D.G. (2009). Tensions between language and discourse in north american knowledge organization. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 10-16. Available: http://iskocus.org/nasko2009-proceedings.php.

CAMPBELL, D.G.; BRUNDIN, M.; MACLEAN, G.; BAIRD, C. (2007) Everything old is new again: Finding a place for knowledge structures in a satisficing world. In Tennis, J.T., Ed. *Proceedings North American Symposium on Knowledge Organization*, 1. Toronto, pp. 21-30. Available: http://iskocus.org/nasko2007-proceedings.php.

DAHLBERG, I. Knowledge organization: its scope and possibilities. Knowledge Organization, [S.l.]: v.20, n.4, p.211-222, 1993.

DAHLBERG, I. (1993). Knowledge organization: its scope and possibilities. *Knowledge organization* 20: 211-22.

DAHLBERG, I. (2006) Knowledge organization: a new science? *Knowledge organization* 33: 11-2.

DAHLBERG, I. Teoria do conceito. Ciência da Informação, Rio de Janeiro: v.7, n. 2, p. 101-107, 1978.

DAHLBERG, I. The Information Coding Classification (ICC): A modern theory-basead fully-faceted, universal system of knowledge field. Axiomathes. [S.l.]: v.18, n.2, p.161-176, 2008.

DOUSA, T.M. (2009 a). Classical Pragmatism and its Varieties: On a Pluriform Metatheoretical Perspective for Knowledge Organization. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp:1-9. Available: http://iskocus.org/nasko2009-proceedings.php.

DOUSA, T.M. (2009 b). Evolutionary Order in the Classification Theories of C. A. Cutter & E. C. Richardson: Its Nature and Limits. IN: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 76-90. Available: http://iskocus.org/nasko2009-proceedings.php.

DOUSA, T.M. (2011). Concretes, countries, and processes in Julius O. Kaiser's Theory of Systematic Indexing: A case study in the definition of general categories. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 160-173. Available: http://iskocus.org/nasko2011-proceedings.php.

EDWARDS, C. (2011). Soviet classifications and situated knowledges. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 13-22. Available: http://iskocus.org/nasko2011-proceedings.php.

EDWARDS, C. (2013). The Social Role of Public Library Classifications. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

FEINBERG, M. (2007). Beyond retrieval: A proposal to expand the design space of classification. In Tennis, Joseph T., Ed. *Proceedings North American Symposium on Knowledge Organization, 1.* Toronto , pp. 31-43. Available: http://iskocus.org/nasko2007-proceedings.php.

FOX, M. (2011) Protoype theory: an alternative concpt theory for categorizing sex and gender. In: Smiraglia, R. P., ed., *Proceedings from North American Symposium on Knowledge Organization, 3*, Toronto, pp. 13-22.

FOX, M.; REECE, A. (2013). The impossible decision: social tagging and Derridas´ s desconstructed hospotality. In: Proceedings from the *North American Symposium on Knowledge Organization, 4,* Milwaukee, pp. 58-74. Available: http://iskocus.org/nasko2013-proceedings.php.

FRICKÉ, M. (2013). Logic and the organization of information: an introduction. In: *Proceedings from North American Symposium on Knowledge Organization, 4*. Milwaukee: University of Wisconsin-Milwaukee, pp. 75-80. Available: http://iskocus.org/nasko2013-proceedings.php.

FURNER, J. (2009). Interrogating "identity": a philosophical approach to an enduring issue in knowledge organization. *Knowledge organization* 36: 3-16.

GARCIA MARCO, F.J. Avances en Organización del Conocimiento en España. In: Garcia Marco, F.J. (ed.) Organización del conocimiento en sistemas de información y documentación. Getafe: Libr. General, 1997.

GARCÍA MARCO, F.J. (1997). Avances en organización del conocimiento en España. In Organización del conocimiento en sistemas de información y documentación. Zaragoza: Libreria General, pp. 7-12.

GILCHRIST, A. (2011). A note on the theory and practice of knowledge organization and knowledge representation. *Scire,* 17(1): 23-28.

GMÜR, M. (2003). Co-Citation Analysis and the Search for Invisible Colleges: A Methodological Evaluation.Scientometrics, Vol. 57, No. 1, pp. 27-57.

GNOLI, C. (2008). Ten long-term research questions in knowledge organization. *Knowledge organization* 35: 137-49.

GNOLI, C. (2011). Ontological foundations in KO: the theory of integrative levels applied in citation order. *Scire* 17(1): 29-34.

GRAF, A.; SMIRAGLIA, R.P. (2013). Transition in Education: Domain Analysis from the Encyclopedia of Milwaukee. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

GREEN, R. (2009). Melvil Dewey's Ingenious Notational System.In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 91-99. Available: http://iskocus.org/nasko2009-proceedings.php.

GREEN, R.; FALLGREN, N. (2007) Anticipating new media: A faceted classification of material types. In Tennis, J.T., Ed. *Proceedings North American Symposium on Knowledge Organization, 1..* Toronto, pp. 87-99. Available: http://iskocus.org/nasko2007-proceedings.php.

GREEN, R. (2011). See-also relationships in the Dewey Decimal Classification. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 41-51. Available: http://iskocus.org/nasko2011-proceedings.php.

GUIMARÃES, J.A.C. A dimensão conceitual da organização do conhecimento no universo científico da Internacional Society for Knowledge Organization-ISKO: uma análise de domínio com subsídio da garantia literária. Marília: UNESP, 2011. Projeto de Produtividade em Pesquisa (PQ)- CNPq referente ao período 2012-2016.

GUIMARÃES, J.A.C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). Revista Ibero-Americana de Ciência da Informação, -- cidade --, v. 1, n. 1, 2008.

GUIMARÃES, J.A.C. ; SALES, Rodrigo de ; MARTINEZ-ÁVILA, D.; ALENCAR, M. F. The conceptual dimension of knowledge organization in the ISKO proceedings domain: a Bardinian analysis. In: Wieslaw Babik. (Org.). Knowledge organization in the 21st century: between historical patterns and future perspectives. 1ed.Wurzburg: Ergon Verlag: 2014, v. , p. 101-106.

GUIMARÃES, J.A.C.; SALES, R. (2013). Contributions of Julius Otto Kaiser to the theoretical framework of Knowledge Organization: basis for the analytical-synthetic method. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

HJØRLAND, B. Domain analysis: a socio-cognitive orientation for Information Science research. Bulletin of the American Society for Information Science and Technology, [S.l.]: v. 30, n.3, feb./mar. 2004. Disponível em: <http://www.asis.org/Bulletin/Feb-04/hjorland.html>.

HJØRLAND, B. (2008). What is knowledge organization (KO)? *Knowledge organization* 35 (2/3): 86-101.

HJØRLAND, B. (2009). Concept theory. *JASIST*, 60 (8): 1519-36.

HJORLAND, B. Domain analysis in information Science: eleven approaches traditional as well as innovative. Journal of Documentation, v. 58, n. 4, 2002, p. 422-462.

HJØRLAND, B. Meta-analysis should also be visible inside information science (Letter to the Editor). Journal of the American Society for Information Science and Technology, [S.l.]: v.53, n.4 p.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: domain-analysis. Journal of the American Society for Information Science, [S.l.] v. 46, n. 6, p.400-425, 1995.

HOFFMAN, G.L. (2009) Applying the User-Centered Paradigm to Cataloging Standards in Theory and Practice: Problems and Prospects. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 27-34. Available: http://iskocus.org/nasko2009-proceedings.php.

HUDON, M. (2011). Teaching bibliographic classification In The 21st Century. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 63-73. Available: http://iskocus.org/nasko2011-proceedings.php.

HUDON, M. (2013). KO and Classification Education Objectives: Are we keeping up with the transformation of our field? In: Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

JACOB, E. K.; ZHANG, G. (2013). The role of virtual boundaries in knowledge and organization. In: *Proceedings from North American Symposium on Knowledge Organization, 4*. Milwaukee, pp. 127-135. Available: http://iskocus.org/nasko2013-proceedings.php.

JAENECKE, P. (1994). To what end knowledge organization? *Knowledge organization*, 21: 3-11.

KASTEN, J. (2007). Knowledge strategy and its influence on knowledge organization. In: In Tennis, J. T., Ed. *Proceedings of the North American Symposium on Knowledge Organization, 1.* Toronto, pp. 44-54. Available: http://iskocus.org/nasko2007-proceedings.php.

KEMP, R.B. (2007) Classifying marginalized people, focusing on natural disaster survivors. In Tennis, J.T., Ed. *Proceedings North American Symposium on Knowledge Organization, 1.*. Toronto, Ontario, pp. 55-62. Available: http://iskocus.org/nasko2007-proceedings.php.

KIEL, E. (1994). Knowledge organization needs epistemological openness. *Knowledge organization* 21: 148-52

KIPP, M.E.I. (2007). Tagging for health information organisation and retrieval. In Tennis, J. T., Ed. *Proceedings of the North American Symposium on Knowledge Organization, 1.* Toronto, pp. 63-74. Available: http://iskocus.org/nasko2007-proceedings.php.

LA BARRE, K. (2009). The Heritage of Facet Analysis in North America: Past Lessons as Pathways for Contemporary Exploration. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 17-26. Available: http://iskocus.org/nasko2009-proceedings.php.

LOEHRLEIN, A. (2011). An examination of interdisciplinary theory between cognitive categorization and knowledge organization. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 122-129. Available: http://iskocus.org/nasko2011-proceedings.php.

LÓPEZ-HUERTAS, M.J. (2008). Some current research questions in the field of knowledge organization. *Knowledge organization* 35: 113-36.

MACCALL, S.L. (2011). Knowledge Organization under digital inversion: A theory for cooperative librarian organizing practices for online textual artifacts. . In: Smiraglia, R. P., ed., *Proceedings from North American Symposium on Knowledge Organization, 3*, Toronto, pp. 13-22. Available: http://iskocus.org/nasko2011-proceedings.php.

MARCHESE, C.; SMIRAGLIA, R.P. (2013). Boundary Objects: CWA, an HR Firm, and Emergent Vocabulary. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

MARTÍNEZ-ÁVILA, D.; SMIRAGLIA, R. (2013). Revealing Perception: Discourse Analysis in a Phenomenological Framework. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

MCTAVISH, J.; FORTIER, A. (2011). A domain-analytic perspective on exual health in LCSH and RVM. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 83-93. Available: http://iskocus.org/nasko2011-proceedings.php.

MILANI, S. O.; GUIMARÃES, J. A.C. (2011). Biases in knowledge representation: an analysis of the feminine domain in Brazilian indexing languages. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 94-104. Available: http://iskocus.org/nasko2011-proceedings.php.

MOULAISON, H.L.; DYKAS, F.; BUDD, J.M. (2013). The Author and the Person: A Foucauldian Reflection on the Author in Knowledge Organization Systems. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

MOYA-ANEGÓN, F.; HERRERO-SOLANA, V. Análisis de dominio de la revista mexicana de investigación bibliotecológica. Información, cultura y sociedad, [S.l.]: n. 5, 2001, p. 10-28.

OLSON, H. A.; HOWARTH, L.C. (2013). A Space of Transition: Rethinking Surrogates. In: *Proceedings from North American Symposium on Knowledge Organization, 4*. Milwaukee, pp. 161-166. Available: http://iskocus.org/nasko2013-proceedings.php.

PAJEK. Introductory social network analysis with Pajek. School of Infomation University of Michigan. Disponivel em <http://ocw.mit.edu/courses/economics/14-15j-networks-fall-2009/assignments/MIT14_15JF09_pajek.pdf > Acesso em 16 jul. 2014.

PATTUELLI, M.C. (2011). Mapping subjectivity: performing people-centered vocabulary alignment. In: Smiraglia, R. P., ed., *Proceedings from North American Symposium on*

Knowledge Organization-NASKO, v. 3, Toronto, pp. 13-22. Available: http://iskocus.org/nasko2011-proceedings.php.

PIMENTEL, D.M. (2007). Exploring classification as conversation In:). In Tennis, J. T., Ed. *Proceedings of the North American Symposium on Knowledge Organization, 1.* Toronto, pp. 1-8. Available: http://iskocus.org/nasko2007-proceedings.php.

PIMENTEL, D.M. (2011). Towards a typology of classificatory change. . In: Smiraglia, R. P., ed., *Proceedings from North American Symposium on Knowledge Organization, 3*, Toronto, pp. 13-22. Available: http://iskocus.org/nasko2011-proceedings.php.

PIMENTEL, D.M. (2009). The KO roots of Taylor's Value-Added Model. In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 58-67. Available: http://iskocus.org/nasko2009-proceedings.php.

SALES, R. GUIMARÃES, J. A. C. (2013). Contributions of Julius Otto Kaiser to the theoretical framework of Knowledge Organization: basis for the analytical-synthetic method. In: *Proceedings from North American Symposium on Knowledge Organization, 4*. Milwaukee, pp. 167-175. Available: http://iskocus.org/nasko2013-proceedings.php.

SCHATZ, J.; STENNES-SPIDAHL, N.; MILLS, S.; LOEHRLEIN, A. J. (2013). Bibliographic Induction: How KO Systems Optimize Browsing by Supporting Library Users' Prior Knowledge. In: *Proceedings from North American Symposium on Knowledge Organization, 4*. Milwaukee, pp. 176-186. Available: http://iskocus.org/nasko2013-proceedings.php.

SMALL, H. On the shoulders of Robert Merton: towads a normative theory of citation.. Scientometrics, v.60, n.1, p.71-79, 2004.

SMIRAGLIA, R. P. (2007). Performance works: Continuing to comprehend instantiation. In:). In Tennis, J. T., Ed. *Proceedings of the North American Symposium on Knowledge Organization, 1.* Toronto, pp. 75-86. Available: http://iskocus.org/nasko2007-proceedings.php.

Smiraglia, R.P. (2009). Modulation and Specialization in North American Knowledge Organization: Visualizing Pioneers In: In: Jacob, E. K.; Kwasnik, B. (Eds.). (2009). Proceedings from North American Symposium on Knowledge Organization, 2. Syracuse, NY, pp: 35-46. Available: http://iskocus.org/nasko2009-proceedings.php.

SMIRAGLIA, R.P., ed. 2011. Proceedings from North American Symposium on Knowledge Organization,. Vol. 3. Toronto, Canada
SMIRAGLIA, R.P. (2012). Epistemology of domain analysis. In *Cultural frames of knowledge*. Würzburg: Ergon-Verlag, pp. 111-24

SMIRAGLIA, R.P. (2013). Is FRBR A Domain? Domain Analysis Applied to the Literature of The FRBR Family of Conceptual Models. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

SMIRAGLIA, R.P. Domain coherence within Knowledge Organization: People, Interacting Theoretically, Across Geopolitical and Cultural Boundaries." In McKenzie, P.; Johnson, K.;

Stevens, S. (ed.). Exploring Interactions of People, Places and Information, Proceedings of the 39th Annual CAIS/ACSI Conference, University of New Brunswick.

SMIRAGLIA. R. P. , ed. 2012. Epistemology of domain analysis. In: Smiraglia R. P. And HUR-Li Lee, eds Cultural frames of knowledge. Wurburg Ergon Verlag. p. 114

SOERGEL, D. 1971. Dokumentation und Organisation des Wisens: Versuch einer methodischen und theoretischen Grundlegung and Beipel der Sozialwissenschaften [Documentation and Organisation of Knoledge : Towards q Methodological and Theoritical Foundation for the Social Sciences] Berlin, DEU: Dunker & Humbolt; 1971. 380 p. ISBN: 3428024419.

SZOSTAK, R. (2013). Classification for Diversity. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.

TENNIS, J.T. (2008). Epistemology, theory, and methodology in knowledge organization. *Knowledge organization* 35: 102-12.

TENNIS, J.T. (2007) The economic and aesthetic axis of information organization frameworks (extended abstract). In Tennis, J.T., Ed. *Proceedings North American Symposium on Knowledge Organization, 1*. Toronto, pp. 112-113. Available: http://iskocus.org/nasko2007-proceedings.php.

TENNIS, J.T. Two Axes of Domain Analysis. Knowledge Organization, [S.l.]: v. 30, n.3/4, p.191-195, 2003.

THORNTON, K. (2011). Contentious categories. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 217-226. Available: http://iskocus.org/nasko2011-proceedings.php.

TOGNOLI, N. B.; GUIMARÃES, J. A. C.; TENNIS, J. T. (2013). Diplomatics as a methodological perspective for archival knowledge organization. In: *Proceedings from North American Symposium on Knowledge Organization, v. 4*. Milwaukee: University of Wisconsin-Milwaukee, pp. 217-225. Available: http://iskocus.org/nasko2013-proceedings.php.

URBIZAGASTEGUI ALVARADO, R. Elitismo na literatura sobre a produtividade dos autores. Ciencia da Informação, Rio de Janeiro: v.38, n.2, Maio/Ago. 2009

VAN DEN HEUVEL, C. (2011). Multidimensional classifications: Past and future conceptualizations and visualizations. In Smiraglia, Richard P., ed. Proceedings from North American Symposium on Knowledge Organization, 3. Toronto, pp. 105-121.
VANZ, S.A.S.; CAREGNATO, S.E. Estudos de citação: uma ferramenta para entender a comunicação científica. Em questão: revista da Faculdade de Biblioteconomia e Comunicação da UFRGS, Porto Alegre: v. 9, n. 2, p. 295-307, jul./dez. 2003.

ZHANG, L.; KIPP, M.; LEE, H.-L. (2013). Genre Tag … Tag Genre. In Proceedings from North American Symposium on Knowledge Organization, 4. Milwaukee. Available: http://iskocus.org/nasko2013-proceedings.php.
ZINS, C. (2004). Knowledge organization: an epistemological perspective. *Knowledge organization* 31: 49-54.

A relação entre organização do conhecimento e ciência da informação na comunidade científica brasileira: uma investigação no âmbito da ISKO-Brasil

Rodrigo de Sales
Universidade Federal Fluminense
rodrigosales@id.uff.br

1 Introdução

Organizar informação para mais bem recuperá-la parece ter sido o mote da Ciência da Informação desde seu início (BORKO, 1968; SARACEVIC, 1996; SMIT & BARRETO, 2002). Porém, quando o assunto organização da informação se coaduna com o assunto organização do conhecimento, algumas visões distintas surgem e inquietam os estudos da área. Para Guimarães (2008), a organização da informação, enquanto área de estudo que integra a Ciência da Informação, consiste em um dos espaços investigativos dessa ciência, possuindo natureza mediadora na medida em que propicia a interlocução entre os contextos de produção e uso da informação. Para Bräscher e Café (2010), a organização da informação se relaciona com as atividades e processos atinentes à organização material da informação, à organização dos itens informacionais nas unidades de informação.

Nesse sentido a organização da informação se articula fundamentalmente com as práticas desenvolvidas pela Biblioteconomia e pela Ciência da Informação. O mesmo não parece ocorrer com igual tranquilidade quando o assunto é a organização do conhecimento. Segundo Bräscher e Café (2010), a organização do conhecimento diz respeito à organização e à sistematização cognitiva do conhecimento, à organização dos conceitos, bem como à construção de sistemas de organização do conhecimento.

Embora as autoras coloquem a organização do conhecimento no âmbito da Ciência da Informação, assim como o faz a organização dos grupos de trabalhos da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB), outras definições parecem afastar essa relação de pertencimento, onde a organização do conhecimento não figuraria como algo dentro da Ciência da Informação. Talvez, os casos mais explícitos em que a organização do conhecimento não pertença necessariamente à Ciência da Informação sejam encontrados em Dahlberg (1993, 1995, 2006 e 2014) e em Hjorland (2003, 2008), que buscam atribuir à organização do conhecimento certa autonomia, seja enquanto disciplina, enquanto campo de estudo ou, ainda, enquanto nova ciência.

Este estudo pretende explorar esse desprendimento da organização do conhecimento (OC) com relação à Ciência da Informação (CI), ou seja, essa visão que aborda OC e CI como duas áreas separadas.

Pensando na organização do conhecimento desenvolvida pelos pesquisadores brasileiros, o que se investiga por ora é a compreensão de como a comunidade científica brasileira aborda a relação entre a OC e a CI.

Para tanto, empregou-se técnicas de Análise de Conteúdo, definidas por Laurence Bardin, para analisar os trabalhos publicados no âmbito do principal fórum de debate sobre organização do conhecimento no Brasil: o Capítulo Brasileiro da ISKO.

2 Duas perspectivas no âmbito internacional: Dahlberg e Hjorland

A primeira década do século XXI revelou a predominância de uma perspectiva que define a OC como um “fazer” de natureza “operacional” (GARCIA, OLIVEIRA, LUZ, 2000; GREEN, 2002; GÁRCIA GUTIÉRREZ, 2002), cujos objetos de investigação são, principalmente, os conceitos e as estruturas conceituais (KENT, 2000; GREEN, 2002; OHLY, 2008, SMIRAGLIA, 2010), formalizadas instrumentalmente nos sistemas de organização do conhecimento - SOC (KENT, 2000; GREEN, 2002; ZHEREBCHEVSKY, 2010; SOUZA; TUDHOPE e ALMEIDA, 2010). Nesse sentido, percebemos uma OC ligada às atividades atinentes à Ciência da Informação e à Biblioteconomia, mais especificamente ao tratamento temático da informação (FOSKETT, 1973).

Porém, no que se refere à natureza (condição) da organização do conhecimento, Dahlberg (1993, 1995, 2006 e 2014) e Hjorland (2003, 2008), embora com perspectivas distintas, são autores que sustentam, no âmbito da *International Society for Knowledge Organization* (ISKO), um discurso de uma organização do conhecimento enquanto campo de estudo autônomo. Essa ótica de OC enquanto campo de estudo pode ser encontrada, para citar alguns, em Ohly (2012), Guimarães, Oliveira & Gracio (2012) e Barros & Moraes (2012).

Embora distintas, as perspectivas a respeito da OC aqui apresentadas não são necessariamente excludentes e auxiliam na compreensão epistemológica deste espaço científico que, embora "jovem", já não se encontra mais em seu estado embrionário.

No final do século XX o discurso de um novo campo se fundamentava nos enunciados de Dahlberg (1993, 1995), em que a autora definia o escopo, os fazeres profissionais, os aspectos institucionais, a classificação da literatura especializada e as tendências da OC. No início do século XXI, Dahlberg, que continua sendo figura central no discurso que busca consolidar a OC como uma área autônoma e independente, tenta avançar tal perspectiva atribuindo à OC *status* de "disciplina científica" e/ou "ciência".

Em 2006, Dahlberg aborda a organização do conhecimento como um espaço autônomo e mais abrangente que aquele tradicionalmente ligado às bibliotecas e unidades de informação. Desta feita, a autora prefere o termo "disciplina científica", e chega a propor que se encontre uma instituição de trabalho formada por cientistas, organizadores do conhecimento e terminologistas dedicados à coleção, à definição e à sistematização de conceitos de todos os campos de assuntos, por meio de uma estrutura de categorização formalizada por uma Classificação de Codificação de Informação. A autora localiza, dentro de um sistema universal das ciências, a OC como um subcampo da Ciência da Ciência.

Dahlberg (2006) esboça, ainda, a OC como uma possível nova ciência. A autora lança mão de distinções feitas por Alwin Diemer para arriscar tratar a OC como uma nova ciência, com objetos, métodos, ações e conteúdos próprios.

Anos depois, Dahlberg (2014) reafirma as ideias acima mencionadas e enfatiza ser necessário retirar a organização do conhecimento (OC) da Biblioteconomia e da Documentação para acomodá-la dentro da Ciência da Ciência, e, também, para outros

domínios que lidam com insumos taxonômicos, tais como Zoologia, Botânica e Microbiologia. Em 2014, a autora volta a reclamar pela formação de um instituto ou academia para a OC, com cientistas e especialistas engajados em um trabalho conjunto que poderia dar frutíferos resultados para o desenvolvimento da OC e da ciência como um todo. Em ambos os artigos, Dahlberg (2006 e 2014) sequer faz menção à Ciência da Informação (CI), deixando claro sua perspectiva de OC autônoma e independente, completamente desvinculada da CI. Para a autora, a OC definitivamente não deve estar conectada somente às ciências ligadas à informação mas, sim, à ciência ligada a todos os campos do saber - a ciência da ciência.

Também no início do século XXI, Hjorland (2003, 2008) trouxe ao debate da OC a distinção entre a organização cognitiva do conhecimento e a organização social do conhecimento. Relativo à perspectiva cognitiva, Hjorland (2003, 2008) afirmava se tratar da organização do conhecimento abordada especialmente pela Biblioteconomia e pela Ciência da Informação, cujo foco recaía nos estudos atinentes aos processos e às construções instrumentais de OC, tais como descrição de documentos, classificação e indexação.

Por organização social do conhecimento, Hjorland (2003, 2008) afirma se tratar basicamente das organizações e categorizações das profissões e das disciplinas, ou seja, a divisão social do labor mental, como por exemplo, a organização dos currículos universitários e também as classificações de assuntos que dividem e relacionam as diferentes disciplinas, bem como as tabelas oficiais que formalizam as ocupações profissionais de um país.

No que diz respeito à interdisciplinaridade da OC, a exemplo de Dahlberg (1993), que propôs a aproximação da OC com as áreas sociais (educação, política, indústria e sociologia), afirmando se tratar de um campo interdisciplinar que se valia, pelo menos, da Filosofia e da Terminologia, Hjorland (2008) ressalta não ser a OC um assunto exclusivo da Biblioteconomia e da Ciência da Informação, mas também de domínios como a Computação, a Linguística e o Processamento de Linguagem Natural, a Teoria do Conhecimento, a Teoria da Organização Social e a Metafísica/Ontologia.

Pode-se afirmar estarmos longe de um consenso a respeito da efetiva ligação epistemológica entre a CI e a OC. Ora encontramos uma tradição, própria da

Biblioteconomia e da Ciência da Informação, que prefere abordar a OC como fundamentalmente discutida no âmbito do tratamento temático da informação, que se caracterizaria como um espaço investigativo pertencente à CI, respaldada, por exemplo, pelos estudos de Foskett (1973); ora encontramos uma perspectiva que desmonta essa relação de pertencimento, alegando que a CI seria, junto à Biblioteconomia, disciplina central que serviria de aporte para conduzir os assuntos de parte da OC (notadamente, a organização cognitiva do conhecimento) – Hjorland. Tem-se, ainda, uma ótica que desvincula definitivamente a OC da CI – Dahlberg.

3 A pesquisa de OC no contexto científico brasileiro

No Brasil, os estudos de OC inicialmente encontraram lugar nas pesquisas desenvolvidas pelo Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), na década de 1970. Porém, foi no âmbito do Grupo de Trabalho *GT 2 – Organização e Representação do Conhecimento*, da ANCIB, que pesquisadores brasileiros puderam alavancar e fortalecer o desenvolvimento da OC no país. Os grupos de trabalhos da ANCIB representam grandes temáticas especializadas estudadas na área da Ciência da Informação. Nesse contexto, a OC, figurava como tema, ou subcampo, da Ciência da Informação.

Nas duas edições do Capítulo Brasileiro da ISKO (2011 e 2013), foram apresentados estudos de pesquisadores brasileiros e estrangeiros dedicados ao desenvolvimento teórico e prático da OC. O Capítulo Brasileiro da ISKO vem se consolidando como o principal fórum de debates relativos à OC no país.

Evidentemente que a pesquisa de OC no Brasil está fundamentada também na literatura internacional de *Knowledge* Organization que, conforme apresentado, revelam, ao menos, três perspectivas distintas no que se refere sua relação com a CI:

Perspectiva 1: evidencia a intenção de independência enquanto disciplina científica, posicionando-se como um subcampo de uma Ciência da Ciência. Quando Dahlberg procura definir o objeto, as atividades, os métodos, as metas dos pesquisadores, a classificação dos assuntos tratados, a criação de uma sociedade internacional, bem como quando lança mão de termos como “organizador do conhecimento”, ao se referir

aos pesquisadores da OC, “campo de estudo” ou “disciplina científica”, ao se referir ao espaço pretendido pela OC, fornece indícios da intencionalidade de tratar a organização do conhecimento como campo ou disciplina independente da CI;

Perspectiva 2: vale-se de assuntos próprios da CI (especialmente os relacionados à recuperação da informação e às tecnologias de informação) para resolver apenas parte da OC, mais especificamente, a parte relativa à organização cognitiva do conhecimento, como aqueles assuntos que dizem respeito às unidades de informação. Essa perspectiva coloca a CI como parte integrante da OC, que, por sua vez, preocupa-se com questões mais abrangentes. Hjorland, principal incentivador dessa perspectiva, parece querer desmontar a relação de pertencimento em que a OC figuraria como um tema ou um subcampo da CI;

***Perspectiva 3*:** tradicionalmente coloca a OC como parte integrante da CI e parece não buscar independência, mas contribuir para o espaço central da CI. Essa ótica encontra apoio na tradição que coloca a OC em diálogo direto com a organização temática da informação, desenvolvida no e para o ambiente informacional. A organização estrutural dos grupos de trabalhos da ANCIB corrobora com essa perspectiva, uma vez que a OC figura como um dos grupos de trabalho inseridos na CI.

Nesta pesquisa, utilizou-se essas três perspectivas como base para nortear a investigação que buscou compreender como os pesquisadores da ISKO-Brasil vêm relacionando a OC e a CI.

4 Aplicação da análise e apresentação dos resultados

O escopo para a formação do corpus de análise foi definido por todos os textos publicados por pesquisadores brasileiros nos Anais da ISKO-Brasil até hoje, 2011 e 2013. Portanto, em um universo de 86 publicações, contando que 19 delas são de autoria estrangeira, o corpus de análise ficou definido em 67 publicações de autoria brasileira. Cabe ressaltar que, embora todas as 67 publicações tenham sido analisadas, nem todas fazem menção aos aspectos relativos à natureza ou à relação da OC com a CI, não figurando, assim, nos resultados da análise. Entraram definitivamente na análise apresentada a seguir somente os textos que em algum momento se referem à natureza

e/ou à relação da OC com a CI, que totalizaram *20 textos*. Ou seja, dos 67 textos que foram analisados, somente 20 (29,8%) mencionaram alguma informação a respeito da natureza da OC e/ou a respeito da relação entre OC e CI.

Como medida de classificação dos dados levantados, Bardin (2003) sugere o processo de *categorização*, que consiste no agrupamento em classes dos elementos convergentes em suas características. As categorias definidas foram: ***Natureza da OC*** e ***Relação entre OC e CI*.**

Para organizar a análise das informações extraídas dos textos, cada categoria foi orientada por variáveis de inferência, que permitiram investigar mais a fundo as ideias apresentadas pelos autores dos textos.

Seguindo o objetivo deste trabalho, foi necessário, primeiramente observar como os autores definem a natureza da OC. Assim, as variáveis de inferências para a categoria ***Natureza da OC*** foram: 1 - território científico; 2 - ação; 3 - assunto e; 4 - ciência.

Isto é, quantos e quais textos consideram que a OC é um *território científico*, como, por exemplo, um campo, uma disciplina, um espaço investigativo, um domínio ou uma área de estudo. A junção de todos esses termos em uma mesma variável não significa considerá-los como de conceitos iguais, mas, tão somente, uma técnica de agrupamento de concepções que apresentam características semelhantes. Neste primeiro caso, respeitando os termos utilizados pelos próprios autores, que ora chamam a OC de campo de estudo, ora chamam de disciplina ou, ainda, de domínio, procurou-se preservar aquilo que há de comum entre os termos adotados, ou seja, a questão territorial – a OC como um território científico (seja ele físico ou abstrato). O mesmo se aplica à definição das demais variáveis: *ação* – a OC vista como um processo, um conjunto de práticas ou, ainda, como uma atividade; *assunto* – para aqueles que a considera como um tema, ou uma linha de pesquisa ou, ainda, como um objeto de estudo. A última das variáveis desta categoria parece ser a mais óbvia, *ciência*, para aqueles que consideram a OC como uma nova ciência.

O próximo passo foi definir as variáveis que permitiriam mais bem organizar e compreender as informações relativas a segunda categoria, *Relação entre OC e CI: perspectiva 1 - a OC como um território científico* ***autônomo e sem relação com a CI***;

perspectiva 2 - a OC como um território científico ***autônomo mas que se relaciona com a CI;***

perspectiva 3 - a OC como ***tema próprio da CI****, ou seja, a* ***OC como um subcampo da CI.***

Relativo à primeira categoria (Gráfico 1), é notório o fato de que a maioria dos textos, 50% (10 textos), consideram que a OC é um território científico, seja ela um campo, uma disciplina, um domínio ou uma área de estudos. Isto evidencia que na realidade brasileira a OC vem alcançado contornos de espaço investigativo próprio, transbordando a concepção de uma dimensão procedimental que a considera como uma ação, 15% (3 textos), e de uma dimensão temática, que a considera como um assunto 15% (3 textos). Cabe destacar que a concepção alavancada por Dahlberg no cenário internacional, que especula a OC como uma nova ciência, não parece ecoar significativamente no contexto da ISKO-Brasil, pois, tal concepção foi retratada em somente 1 dos textos analisados (5%).

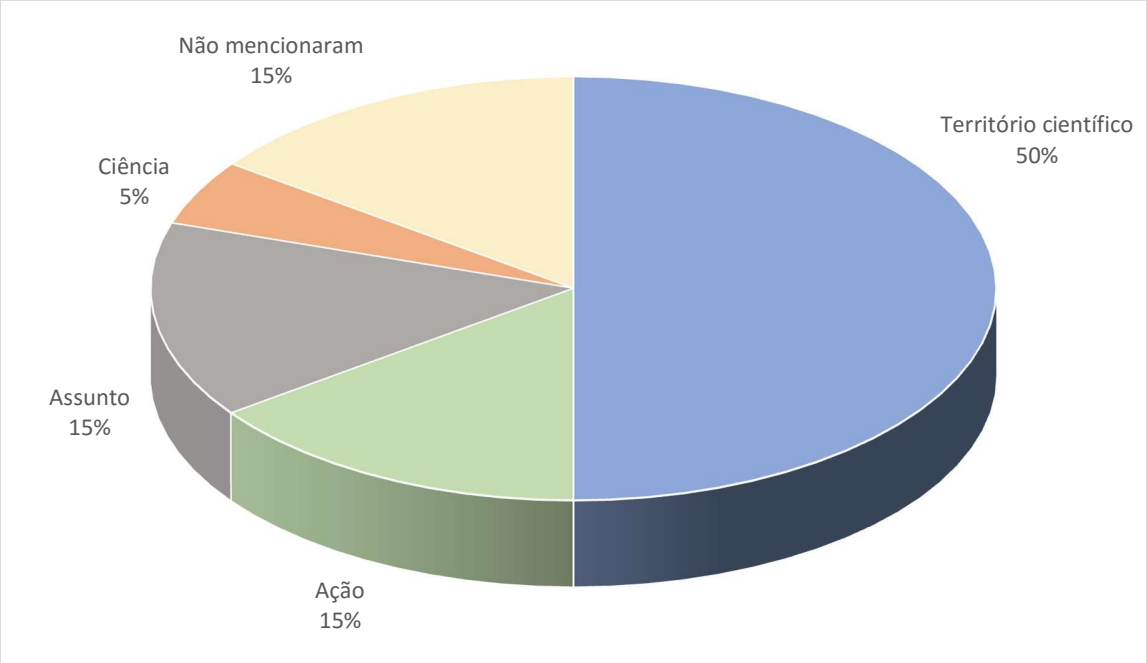


Gráfico 1: Natureza da OC (Fonte: Elaborado pelo autor)

No que diz respeito à relação entre a OC e a CI (Gráfico 2), ponto de observação principal deste estudo (contemplado na categoria 2), observa-se que a maioria dos textos analisados, 35% (7 textos), convergem com a perspectiva 2, entendendo que a OC é um

espaço investigativo autônomo mas que guarda forte interlocução com a CI. Logo em seguida, 4 textos (20%) entendem que a OC é de fato um tema especializado ou uma subárea da CI, enquadrando-se na perspectiva 3. Somente 1 dos textos do corpus, 5%, converge com a perspectiva 1, de que a OC é um território autônomo que não se relaciona com a CI, ou, como prefere Dahlberg, que a OC é uma nova ciência.

Vale destacar os percentuais dos textos que não fizeram menção alguma a respeito da natureza da OC (Gráfico 1) e da relação entre OC e CI (Gráfico 2). Enquanto 15% dos textos analisados não mencionaram informações a respeito da natureza da OC, 40% não o fizeram a respeito da relação entre OC e CI. Este dado revela como estudos desta natureza são cada vez mais necessários para uma melhor compreensão epistemológica da OC.

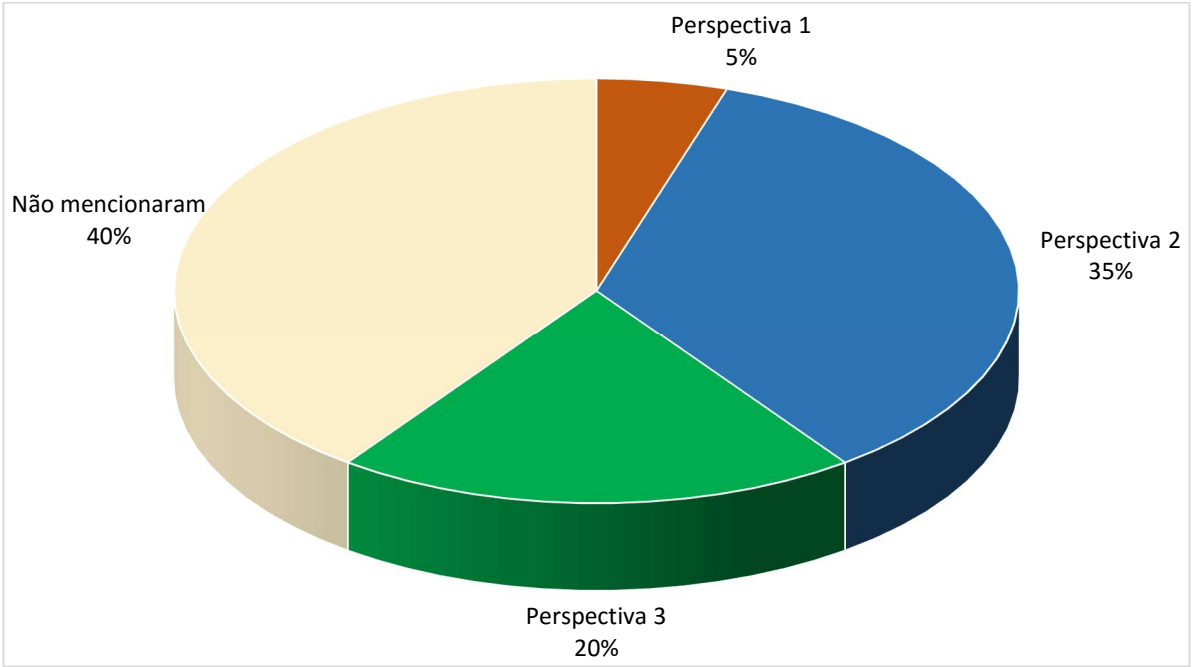


Gráfico 2: Relação entre OC e CI (Fonte: Elaborado pelo autor)

Com isso, o cenário que se descortina no âmbito da ISKO-Brasil é o de que os pesquisadores brasileiros, em sua maioria, entendem que a OC alcançou um status de território científico autônomo que está em constante diálogo com a Ciência da Informação. Em outras palavras, a OC, segundo esta comunidade científica, não se configura mais apenas como o tema pertencente à Ciência da Informação. O protagonismo da OC como subcampo da CI, como abordado pela própria pesquisa

brasileira na virada do século passado para o século XXI, no que se refere à sua condição científica, é dividido agora com a ideia de que é OC consiste em um novo território que, transbordando a perspectiva inicialmente desenvolvida no país, imprime seus próprios contornos epistemológicos e se revela como um espaço emergente.

5 Conclusão

O fato da CI ser, na virada do século XX para o XXI, a área responsável por abarcar e desenvolver as pesquisas atinentes à OC no Brasil não impediu que esta última transbordasse os limites daquela e criasse gradativamente seus próprios contornos. A institucionalização da OC por meio do Capítulo Brasileiro da ISKO, pode ter sido tanto a causa quanto a consequência desse transbordamento na comunidade científica brasileira.

Porém, como mostrou a pesquisa, esse novo delineamento, que galga em busca de sedimentação epistemológica e que já marca um novo espaço, não significa um descolamento absoluto da OC com a CI. O vínculo que a maioria dos pesquisadores brasileiros de OC têm com a CI, seja de ordem institucional, seja de ordem epistemológica, reflete um pensamento majoritário que prefere entender ambas as áreas como espaços de imprescindível diálogo e interseção, não cabendo mais aquela relação de pertencimento de outrora.

Referências

BARDIN, L. *L' analyse du contenu*. 7ème. Paris: PUF, 2003. 296 p. (Le Psychologue, 69).

BARROS, T. H. B.; MORAES, J. B. E. Archival classification and knowledge organization: Theoretical possibilities for the archival field. In *Categories, contexts and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference 6-9 August 2012 Mysore, India*. Advances in Knowledge Organization, 13. Würzburg: Ergon, 2012, p. 272-276.

BORKO, H. Information science: what is it? *American Documentation*, v.19, n.1, p.3-5, Jan. 1968.

BRASCHER, M; CAFÉ, L. Organização da informação ou organização do conhecimento? In: LARA, M. L. G.; SMIT, J. (Org.). *Temas de pesquisa em Ciência da Informação no Brasil.* São Paulo: Escola de Comunicação e Artes/USP, 2010. ISBN: 978-85-7205-081-4.

DAHLBERG , I. Current trends in Knowledge organization. In: Garcia Marco F. J. (Org.). *Organización del conocimiento em sistemas de información y documentación*. Zaragoza: Universidad de Zaragoza. 1995, p. 7-25.

DAHLBERG, I. Knowledge organization: a new science? *Knowl. Org.* v. 33, n. 1, 2006, p. 11-19.

DAHLBERG, I. Knowledge organization: its scope and possibilities. *Knowl. Org.* v. 20, n. 4, 1993, p. 211-222.

DAHLBERG, I. What is knowledge organization. *Knowl. Org,* v. 40, n. 1, 2014, p. 85-91.
FOSKETT, A.C. *A abordagem temática da informação*. Tradução de Antônio Agenor Briquet de Lemos. São Paulo: Polígono; Brasília: Ed.UnB, 1973.

GARCIA, S. M. M.; OLIVEIRA; LUZ, G. M. S. Knowledge organization for query elaboration and support for technical response by the internet In: *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference* Würzburg: Ergon. 2000. p.189.

GÁRCIA GUTIÉRREZ, A. L. Knowledge organization from a culture of the border: towards a trascultural ethics of mediation. In: *Challenges in knowledge representation and organization for the 21st century: integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference*. Würzburg: Ergon, 2002, p.518.

GREEN, R. Conceptual universals in knowledge organization and representation In: Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: *Proceedings of the Seventh International ISKO Conference*. Würzburg: Ergon, 2002, p.15.

KENT, R. E. The information flow foundation for conceptual knowledge organization In: *Dynamism and stability in knowledge organization: Proceedings of the Sixth International ISKO Conference*. Würzburg: Ergon. 2002, p. 111.

GUIMARAES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-Americana de Ciência da Informação*, v. 1, p. 77-99, 2008.

GUIMARÃES, J. A. C.; OLIVEIRA, E. T.; GRACIO, M. C. C. Theoretical referents in knowledge organization: A domain analysis of the knowledge organization journal. In

Categories, contexts and relations in knowledge organization: Proceedings of the Twelfth International ISKO Conference 6-9 August 2012 Mysore, India. Advances in Knowledge Organization, 13. Würzburg: Ergon. 2012, p. 31-38.

HJORLAND, B. Fundamentals of knowledge organization. *Knowl. Org.* v. 30, n. 2, p. 87-111, 2003.

HJORLAND, B. What is knowledge organization (KO)? *Knowl. Org.* v. 35, n. 3/2, p. 86-111, 2008.

OHLY , H. P. Knowledge organization pro retrospective In: *Culture and identity in knowledge organization: Proceedings of the Tenth International ISKO Conference*. Würzburg: Ergon, 2008, p. 210.

OHLY , H. P. Mission, programs, and challenges of knowledge organization. In: *Categories, contexts and relations in knowledfe organization: Proceedings of the Twelfth International ISKO Conference*. Mysore: Ergon, 2012, p. 15-23.
SARACEVIC, T. Ciência da Informação: origem, evolução e relações. *Perspectivas em Ciência da Informação,* Belo Horizonte, v. 1, n. 1, jan./jun. 1996, p. 42-62.

SMIRAGLIA, R. P. Perception, knowledge organization and noetic affective social tagging. In: In: *Paradigms and conceptual systems in knowledge organization:* Würzburg: Ergon, 2010, p.64.

SMIT, J. W.; BARRETO, A. A. Ciência da Informação: base conceitual para a formação profissional. In: VALENTIN, M. L. P. (Org.). *Formação do profissional da informação*. São Paulo: Polis, 2002. Cap. 1, p. 9-24.

SOUZA, R. R.; TUDHOPE, D.; ALMEIDA, M. B. The KOS spectra: a tentative faceted typology of knowledge organization systems. In: *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference*. Würzburg: Ergon, 2010, p.122.

ZHEREBCHEVSKY. Formalism in knowledge organization. In: *Paradigms and conceptual systems in knowledge organization: Proceedings of the Eleventh International ISKO Conference*. Würzburg: Ergon, 2010, p. 98.

Poole, o índice e as fraturas: Indexação e publicações seriadas nos Estados Unidos do Oitocentos

Gustavo Silva Saldanha
Instituto Brasileiro de Informação em Ciência e Tecnologia
Universidade Federal do Estado do Rio de Janeiro
Janeirosaldanhaquim@gmail.com

Naira Christofoletti Silveira
Universidade Federal do Estado do Rio de Janeiro
naira.silveira@unirio.br

1 Introdução à Fratura: da e para uma racionalidade bibliográfica

> La Bibliographie des Périodiques et des Journaux est une oeuvre immense, elle ne peut être réalisée que par cooperation et publiée sous forme de recueils divers. Elle est appelée à être concentrée dans le *Répertoire Bibiographique Universel.* (OTLET, 1934)

Entre os verbetes *imprimeur* e *insectes* do *Dictionnaire raisonné de bibliologie*, de Gabriel Peignot (1802), a ausência de um termo no desdobrar da ordem alfabética sugere uma grande questão epistêmico-histórica para os estudos informacionais: o desenvolvimento das práticas de indexação (em seu sentido inaugural de arte de criar índices) parece não ocupar lugar central no pensamento bibliológico reflexivo que ali nascia. Considerada a exaustiva revisão peignotiana, reconhece-se a existência da discussão sobre os índices, mas não com o destaque de um conceito que merece a singularidade de um verbete. Nem *index,* muito menos, naturalmente, o conceito de *indexation,* comparecem em destaque na compilação peignotiana do início do Oitocentos. De fato, o discurso do bibliotecário francês orienta-se para um ponto de vista prospectivo – uma ciência por nascer –, mas tem no século XVIII sua vasta fonte. O século seguinte, "aberto" pelo *Dictionnaire*, por sua vez, verá o rápido desenvolvimento não apenas das

práticas de indexação, como também da reflexão sobre tal conceituação diante do dilema da multiplicação de artefatos textuais e de línguas de especialidade.

O século XIX pode ser representado, dentre outras características, pela corrida em busca de possibilidades de constituição de uma língua científica – um dos limites epistemológicos do positivismo comteano – que permitiria a identificação e a construção de novos conhecimentos empíricos. Acompanha este percurso o desenvolvimento de teorias, de abordagens e de experimentos voltados à elaboração de formas de registro e de representação dos registros da linguagem, traduzidos como registro do conhecimento (índice objetivo de potencialização da "naturalização" das relações entre conhecimento, linguagem e materialidade ao longo da Modernidade).

O que tratamos por "fratura" no presente documento parte inicialmente de duas noções correlatas para nosso propósito. A ideia remete aos sentidos de rompimento, de quebra, de ruptura, de fenda. Ela ilustra, nesta reflexão, este desdobrar de especializações junto de uma fragmentação da filosofia ou de um saber geral, doador de formas e conteúdos aos demais. No vocabulário bibliográfico, "fratura" nos remete ainda a outro significado preponderante para esta análise: trata-se da escrita gótica, caracterizada pelo "contraste entre traços grossos e finos", que, tendo nascido no século XI, provocou grande influência no séculos XII e XIV, chegando ao uso ainda no século XX, na Alemanha (FARIA, PERICÃO, 2008, p. 303).

A metáfora da acepção técnica de "fratura" é complementar e mais objetiva ao nosso discurso: temos aqui uma transformação técnica na racionalidade bibliográfica exatamente no processo histórico que vai do Renascimento Cultural ao Renascimento Artístico, fundando uma visão de mundo nos primeiros cinco séculos do segundo milênio de nossa datação ocidental devotada à empiria, aos marcos indutivos e a um "modo documental" de se fazer ciência. Junto das diversas especializações científicas, estabelece-se a demanda por uma disciplina específica, por uma teoria do conhecimento, de caráter metacientífico, ou, nas visões de Peignot (1802) e Otlet (1934), estamos diante de outro resultado da "fratura", o aparecimento (ou a demanda de construção) da Bibliologia. Suplementar e não menos importante, outro sentido, oriundo do contexto de uma teoria do conhecimento bibliográfico, aqui se reconhece: as fragmentações do Livro (como a visão otletiana indicava, em letra maiúscula) em produções seriadas e, nestas,

a avalanche de incontáveis cadeias de novos discursos, saberes, línguas de especialidade.

Para solucionar o problema de uma multiplicação contígua de linguagens e novos instrumentos, ou seja, junto das novas “comunidades discursivas” que interagem em determinados “domínios” (HJORLAND, ALBRETSCHEN, 2005) no curso do século XIX, as intervenções teóricas e práticas dos artífices dedicados à organização dos saberes se apresenta. É necessária, diante da miríade de linguagens, uma gama de instrumentos para metarrepresentar os artefatos simbólicos construídos por cada comunidade dispersa na paisagem e fechada em seus léxicos, seus instrumentos, seus métodos, como índices, bibliografias e catálogos.

Vinculado ao contexto espaço-temporal retratado, reconhecemos aqui a presença de William Frederick Poole, como uma figura central para a consolidação de um pensamento americano em recuperação temática de artefatos, principalmente no âmbito dos periódicos, com influência mundial. Diante da dinâmica da produção seriada e da profusão de conteúdos que desta pode advir, percebe-se aqui uma experiência sistemática, fruto de um diálogo entre teoria e prática, para a representação temática dos instrumentos de memória e de comunicação de comunidades discursivas distintas que se multiplicavam diante das “fraturas do conhecimento”. No âmbito de uma epistemológica histórica, interessada no reconhecimento e na discussão de conceitos, de métodos, de abordagens e de experimentos no espaço-tempo discursivo dos estudos informacionais, o objetivo do trabalho é discutir, a partir de um enfoque teórico, o desenvolvimento de ferramentas de recuperação da informação no âmbito da produção seriada no século XIX, tendo como Poole como foco analítico deste contexto.

2 As Fraturas Tratadas: da construção de índices à indexação

> The only comprehensive general index to periodical literature is that prepared by Mr. Poole [...] (ROBINSON,1876, p. 665)

Ao discorrer sobre a bibliografia de periódicos e de jornais, Paul Otlet (1934) destacava os desafios inerentes à prática de representação descritiva e temática do artefato, ou seja, sua complexidade e a demanda não apenas por teorias e métodos para sua apropriação adequada, como também políticas nacionais e internacionais para sua produção e seu uso. Questões como a normalização de abreviaturas de títulos de periódicos, dados dos editores, registros de interrupção, indicação de bibliotecas que guardam e disponibilizam tais documentos representam algumas destas demandas.

Dentre os trabalhos clássicos dedicados a esta preocupação, Otlet (1934) sublinhava o *Schedularium* do Concilium Bibliographicum e o *Catalogue of Scientific Papers*, do século XIX, além do *Catalogue Internacional de la Littérature Scientifique*, este último do século XX, desenvolvido entre 1900 e 1914. Nesta relação, o advogado belga destaca que "Le *Poole Index* a été la première oeuvre de cette nature" (OTLET, 1934, p. 296). É curiosa a menção, uma vez que se trata de um índice americano, reconhecido por um teórico do Velho Mundo como uma vanguarda na preocupação com a compreensão da dinâmica do conhecimento recaracterizado em fraturas disciplinares. Tal preocupação é o indício claro da relação política entre o desenvolvimento de uma indexação nascitura em território norte-americano, a produção intensa de conhecimento registrado e uma demanda pelo acesso ao conhecimento fragmentado.

O chamado Índice Poole para a literatura periódica procurava identificar, relacionar e representar os conteúdos de documentos periódicos de diversa natureza, visando a sua recuperação. Foi, portanto, para alguns autores, o "[...] primeiro instrumento de pesquisa desse gênero" (BATTLES, 2003, p. 142). Segundo Otlet (1934), a ferramenta pooleiana chegou a registrar 590.000 artigos de 470 periódicos diferentes e foi continuado pelo *Annual Literary Index* e pelo *Annual Library Index.*

Em razão de suas características, tornou-se um instrumento muito relevante para o desenvolvimento da indexação e está presente até os dias atuais. O fato é corroborado por Silva e Fujita (2004, p. 140), que relatam que o índice publicado em 1882

> [...] introduziu os índices cumulativos de periódicos, usando para artigos de jornais entradas de assunto representados pelas palavras-chave retiradas dos títulos dos artigos indexados [...] W.F. Poole foi responsável pelos grandes índices dos periódicos do século XIX.

Mesmo depois de tantos anos, observa-se que os motivos que impulsionaram Poole para a construção de seu índice permanecem inalterados, fornecer um instrumento para recuperar as informações dispersas. Conforme constam nas definições que se seguem:

> **INDEXES**: Indexes provide subject, author, and/or title indexing to a particular set of periodicals and gives a full citation for each article. The citation includes the title of the periodical, date, volume, pages, as well as the author and title of the article. Some indexes include ABSTRACTS. (INDEXES, 2014, *online*)
>
> index
> Points to where information can be found.
> List at the end of books, encyclopedias, etc. that indicates by author, title and/or subject the location of information within the book or encyclopedia. Tool that arranges (by author, title, or subject) citations to articles in a selected group of periodicals. (INDEX, 2014, *online*)

Embora sua origem esteja relacionada aos documentos seriados, especificamente aos periódicos, atualmente é comum encontrarmos índices em vários tipos de documentos ou que rementam a mais de um documento. Os periódicos podem ser considerados como documentos contínuos. Muitas vezes, são publicados por muitos anos e cada volume pode abarcar uma série de artigos com muitos assuntos dispersos em seus números. Sendo assim, se define "**publicação periódica:** Publicação em qualquer tipo de suporte, editada em unidades físicas sucessivas, com designações numéricas e/ou cronológicas e destinada a ser continuada indefinidamente" (ABNT, 2002, p. 2).

Devido às características dos períodicos, é fácil compreender porque este tipo de documento motivou a criação de índices. Alguns periódicos publicam periodicamentes seus índices com a finalidade de recuperar o conhecimento disperso e fragmentado em sua coleção. A elaboração de índices está relacionada à recuperação de informações dispersas, por isso, muitas vezes esse instrument está relacionado a Documentação, embora ele tenha surgido em um ambiente Biblioteconômico. Para Lancaster (1993, p. 1), "O propósito principal da elaboração de índices e resumos é construer representações de documentos publicados numa forma que se preste a sua inclusão em algum tipo de

base de dados". À época de Poole, essa base de dados era um documento impresso, elaborado com a melhor tecnologia daquele momento.

Os índices são compostos por termos extraídos de um documento mediante ao processo de indexação. Tais termos tranformam-se em descritores de um conteúdo que serão utilizados na busca e recuperação da informação (LANCASTER, 1993).

O processo de indexação está atrelado à representação de assuntos,

> *Indexação de assuntos* é uma expressão usada de modo mais impreciso; refere-se à representação de conteúdos temáicos de partes de itens bibliográficos completes, como é o caso do índice no final de um livro. […] O processo pelo qual o conteúdo temático de itens bibliográficos é representado em bases de dados publicadas – em forma impressa ou legível por computador – é quase invariavelmente designado como *indexação de assuntos",* quer se estejam examinando itens global ou parcialmente (LANCASTER, 2003, p. 15).

As definições acima reconhecidas compõem hoje o quadro vasto da teoria da indexação. Esta, junto de seus usos, seus experimentos, seus instrumentos, é causa e fruto de transformações sócio-políticas, que tiveram no século XIX um avanço acelerado, acompanhando a multiplicação de artefatos de ordem seriada. Neste sentido, o pensamento e o desdobrar empírico das ações de Poole são um modelo elementar para a compreensão da construção desta teoria e seus conceitos.

3 Poole e seu Índice no Espaço-Tempo: indexação, política e a construção da *Library Science* no Oitocentos

> "... and we recommend that a similar system of indexing be extended to the transactions and memoirs of learned societies." (CLAPP, 1954)
> [Mensagem da Conferência de Bibliotecas nos Estados Unidos em 1853]

Se o pensamento biblioteconômico estadunidense repercutirá a partir do final do século XIX no mundo, estabelecendo-se como grande fonte de influência para o desenvolvimento do que ele próprio passou a chamar futuramente de *Library and Information Science*, suas abordagens teóricas nascituras do Oitocentos têm impacto epistemológico direto nesta constituição. No entanto, a epistemologia não se encontra

jamais em estado “puro”, em uma condição de “isolamento social” e “a-político”. Os índices têm um papel fundacional (junto de outras abordagens) no pensamento americano biblioteconômico.

Em 1848, Poole, iniciando sua carreira como bibliotecário na Universidade de Yale, desenvolve o estudo dos periódicos da biblioteca universitária, procurando, ao invés de realizar listas tradicionais, compreender os temas envolvidos nos artigos diversos presentes nos artefatos seriados. No mesmo ano, o manuscrito de Poole chama a atenção e é publicado por George P. Putnam, “*An Alphabetical Index to Subjects Treated in the Reviews, and Other Periodicals, to Which No Indexes Haw Been Published: Prepared for the Library of the Brothers in Unity, Yale College.”* O trabalho de Poole é um desenvolvimento sistemático dos estudos de John Edmands, bibliotecário anterior de Yale, desenvolvido em 1947. (CLAPP, 1954)

Após o primeiro “impacto”, apenas no “miraculoso” ano de 1876, é retomada a proposta. Agora, o foco central estava no desenvolvimento cooperativo do índice, contando com 51 bibliotecas, com foco na descrição de 232 periódicos. A recepção do trabalho, segundo Clapp (1954), é claramente grandiosa: o Museu Britânico é uma das primeiras instituições a adquirir a obra e preocupar-se com seu desenvolvimento. O contexto das últimas três décadas do Oitocentos nos Estados Unidos demarca já uma espécie de “indústria de indexação”, ou um conjunto de indivíduos, instituições públicas e corporações privadas dedicadas ao “mercado das metarrepresentações de fundo temático”.

A partir do recebimento dos dados de cada biblioteca acerca de suas revistas, o bibliotecário redistribuiria as metarrepresentações para todas as cooperantes.

> O projeto, em poucas palavras, foi este: eu imprimiria e enviaria para as principais bibliotecas uma lista de periódicos que deveriam constar no índice. As bibliotecas verificariam a presença dos periódicos e retornariam com as listas particulares. Tendo recebido estas listas, eu faria uma distribuição equitativa dos trabalhos, tendo uma cota para mim, redistribuindo proporcionalmente os itens para as bibliotecas maiores e menores. Cada biblioteca elaboraria seu índice de acordo com um código de regas fornecidas e enviaria as referências para mim, que iria revisar, organizar e incluir. (POOLE, 1882, tradução nossa)

O trabalho cooperativo repercutiu no Reino Unido e as trocas se estabeleceram também além Atlântico. Poole (1882) destaca, na discussão teórico-técnica de

elaboração do índice, diretrizes como padronização, abrangência, qualidade dos descritores, abreviações, datação. O indexador também chama a atenção para o propósito geral do índice: reconhecendo a demanda das “comunidades discursivas dispersas”, Poole (1882) deixa claro que o propósito geral do Índice é atingir alunos, jornalistas e profissionais do mundo livro, tendo, deste modo, um olhar generalista, sem deixar de reconhecer as línguas de especialidade. Neste sentido, Poole (1882) aponta para experiências de indexação de documentos científicos como das áreas jurídicas e da saúde. As “fraturas do conhecimento” científico reverberam no mundo cotidiano, tendo, por isto os “novos assuntos” sua representação no Índice. Mesmo não sendo o foco, revistas científicas e profissionais eram incluídas quando suas temáticas de aproximavam objetivamente dos elementos generalistas das transformações em curso. O Índice Poole representa, deste modo, uma influência direta para as demandas imediatas de recuperação da informação especializada.

4 Algumas palavras conclusivas

> The "public library" which we are to consider is established by state laws, is supported by local taxation and voluntary gifts, is managed as a public trust, and every citizen of the city or town which maintains it has an equal share in its privileges of reference and circulation. (POOLE, 1876)

No espaço-tempo específico de Poole, percebemos o desenvolvimento de uma relação muito próxima entre recuperação temática de artefatos, democracia e epistemologia da organização dos saberes. Em outras palavras, podemos reconhecer como, de um lado, sob um escopo teórico específico da futura *Library and Information Science* estadunidense, ou seja, a indexação, estabelece-se uma possibilidade de contribuir para o ideal democrático americano do Oitocentos de acesso universal ao conhecimento, o que co-resulta na emergência de um campo específico, uma nova epistemologia (mais um domínio ampliado e emancipado) na fratura da organização geral da árvore do conhecimento, o “fragmento que se dedicará aos fragmentados”.

Referências

CLAPP, Verner W. Indexing and abstracting services of serial literature. *Library trends*, v. 2, p. 509-21, apr. 1954.

BATTLES, M. *A conturbada história das bibliotecas*. São Paulo: Planeta, 2003.

FARIA, M. I. R. de; PERICÃO, M. da G. *Dicionário do livro*: da escrita ao livro eletrônico. São Paulo: Edusp, 2008.

HJORLAND, B.; ALBRECHTSEN, H. Toward a new horizont in information science: domain-analysis. *Journal of the American Society for Information Science*, v. 46, n. 6, p. 400-425, 1995.

INDEX. In: UNIVERSITY OF ILLINOIS. University Library. *Glossary of library terms*. Urbana, IL: University of Illinois, 2014. Disponível em: <http://www.library.illinois.edu/learn/intro/glossary.html#I> . Acesso em: 01 maio 2015.

INDEXES. BOSTON COLLEGE. UNIVERSITY LIBRARIES. *Library Glossary.* Boston: Boston College, 2014. Disponivel em: <http://www.bc.edu/libraries/help/howdoi/howto/glossary.html#i> . Acesso em: 01 maio 2015.

LANCASTER, W. F. *Indexação e resumos:* teoria e prática. Brasília: Briquet de Lemos, 2003.

POOLE, W. F. *Poole's Index to Periodical Literature*. Chicago, 1882. Disponível em: < http://c19index.chadwyck.com/infoCentre/introvol1.jsp>. Acesso em: 25 jan. 2013.

OTLET, Paul. *Traité de documenatation*: le livre sur le livre: théorie et pratique. Bruxelas: Editiones Mundaneum, 1934.

PEIGNOT, G. *Dictionnaire raisonné de bibliologie,* tomo I. Paris: Chez Villier, 1802.

POOLE, W. F. *An Index to Periodical Literature*. New York: C.B. Norton, 1853.

POOLE, W. F. The organization and management of public libraries. In: EUA. Bureau of Education.*Public libraries in United States of America*: their history, condition, and management. Parte I. Whashington: Government Printing Office, 1876. p. 465-476.

ROBINSON, Otis H. On indexing periodical and miscellaneous literature. In: EUA. Bureau of Education. *Public libraries in United States of America*: their history, condition, and management. Parte I. Whashington: Government Printing Office, 1876. p. 663-672.

SILVA, M. dos R. da; FUJITA, M. S. L. A prática de indexação: análise evolutiva de tendências teóricas e metodológicas. *Transinformação*, Campinas, v. 16, n. 2, p. 133-61, 2004. Disponível em: <http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/717>. Acesso em: 01 maio 2015.

Os processos de representação do conhecimento arquivístico: elementos históricos e conceituais da classificação e descrição

Natália Bolfarini Tognoli
Universidade Estadual Paulista
nataliatognoli@marilia.unesp.br

Thiago Henrique Bragato Barros
Universidade Federal do Pará
thiagobarros@ufpa.br

1 Introdução

A Classificação e a Descrição são atividades nucleares no processo de organização e representação de arquivos. Esse conhecimento pode ser entendido como todo conhecimento gerado a partir das atividades realizadas por pessoas físicas ou jurídicas e que resultam em informações orgânicas.

Na Arquivística, o princípio de respeito aos fundos – ou princípio da proveniência, desde a segunda metade do século XIX, é considerado o mais importante nos processos de organização e representação. Estabelecido em 1841, devido à necessidade do governo francês de solucionar os problemas de acumulação desordenada de documentos nos Arquivos Nacionais após a Revolução Francesa, o princípio estabelece que os documentos que são produzidos ou acumulados por uma pessoa ou instituição devem ser agrupados e mantidos juntos, gerando, assim, os chamados *fundos de arquivo*.

Para Duchein (1983, p. 64) “a definição mais simples de *respeito aos fundos* significa agrupar, sem misturar a outros, os arquivos (documentos de todo tipo) criados o recebidos por uma administração, estabelecimento, pessoa ou corporação.

Neste sentido, compreende-se que os documentos reunidos em fundos e\ou coleções refletem o conhecimento produzido sobre uma pessoa ou instituição específica, atuando três principais atores neste contexto: (1) o criador (autor), que é a pessoa física ou jurídica responsável pela criação dos documentos; (2) o usuário, que irá utilizar o

documento para fins de prova, administrativa, legal, social, histórica e\ou científica, e (3) Os gestores, arquivistas, documentalistas, ou seja, os atores responsáveis de algum modo pela organização e socialização destas informações.

Portanto, acredita-se que seja possível compreender o trabalho de representação arquivística como uma forma de organização do conhecimento, ou seja, o arquivista e a Arquivística como área com relação interdisciplinar com a Organização do Conhecimento, definida por Hjørland (2008, p. 88) como “as atividades de descrição, indexação e classificação de documentos realizadas em bibliotecas, bases de dados bibliográficas, e arquivos”.

Apresenta-se, aqui, os elementos históricos e conceituais do processo de representação, com foco nas atividades de classificação e descrição enquanto fundamentais para organização e acesso ao conhecimento gerado a partir das informações orgânicas. Para tanto, seguiremos aqui, a lógica da classificação enquanto a primeira atividade no processo de representação precedendo a descrição.

As atividades de descrição e classificação são, para os arquivos, o que são as atividades de classificação, catalogação e indexação para as bibliotecas, ou seja, a organização do conhecimento arquivístico em si.

Pretende-se com este percurso histórico-conceitual aproximar processos típicos dos arquivos à teoria desenvolvida no espaço teórico-conceitual da Organização do Conhecimento.

2 Classificação Arquivística: Elementos chave para a organização do conhecimento

A Classificação na realidade dos arquivos é considerada uma atividade limítrofe na medida em que constrói uma ponte entre as atividades de gestão e as atividades de acesso e preservação documental.

As atividades de classificação, assim com a descrição, foram as primeiras atividades a serem de algum modo conceituadas na teoria e prática arquivística, desenvolvida a partir do século XIX.

Os princípios básicos da proveniência e ordem original irão estabelecer-se desde o final do século XIX como fundamentais para a organização dos arquivos.

No início do século XX, a classificação dá um pequeno salto com as reconsiderações a respeito do tema publicadas por Hillary Jenkinson, em 1922. Para o autor, a classificação é dividida em duas partes "a primeira estuda a administração da instituição, sua história e organização; a segunda divide os arquivos em classes e suas subdivisões (JENKINSON, 1922, p. 81, tradução nossa)".

A classificação, enquanto teoria, desenvolve-se de maneira plena a partir da década de 1950-1960 com as publicações de Theodore R. Schellenberg, Oliver W. Holmes, Ernest Posner, Petter Scott, dentre outros arquivistas e pesquisadores dos Arquivos Nacionais dos Estados Unidos e da Austrália.

Ao mesmo tempo em que a disciplina avança de maneira profunda, a partir da década de 1950, ocorre uma ruptura que irá perpetuar-se ao longo das décadas até a atualidade.

Trata-se da separação de certas atividades e conceitos determinados pela fase em que os documentos se encontram.

Uma das instituições responsáveis por esse acontecimento é o Arquivo Nacional Americano, centralizado na figura de Theodore R. Schellenberg, que difundiu e popularizou seus métodos, separando a classificação como uma atividade típica de arquivos ativos o arranjo aos arquivos permanentes, dando início de uma ruptura sistemática na disciplina.

Observa-se que anteriormente não existia uma diferença entre essas funções, e o uso mais corrente era do termo arranjo para designar coisas que, para a arquivística americana, depois de década de 1950, são diferentes. Eastwood faz uma ponderação fundamental a respeito desse problema terminológico-conceitual:

> A escolha da palavra “arranjo” como nome deste processo é infeliz. Denota colocar coisas de um modo aceitável, um uma ordem conveniente, e como arranjar livros em uma prateleira. A palavra classificação não é mais satisfatória, neste caso denota o arranjo ou a ordenação de coisas por classes e é um termo melhor reservado na arquivística para o processo de organizar documentos ativos. (EASTWOOD, 2000, p. 93, tradução nossa)

A classificação por funções é hoje, ao mesmo tempo, o trunfo e o simulacro da arquivística, no sentido de que para alguns autores é por esse olhar que se define a atividade como, por exemplo, para Shepherd and Yeo (2003, tradução nossa):

"Esquemas de classificação são baseados na análise das funções, processos e atividades"

A classificação é uma atividade gerencial e de planejamento não só para os arquivos, mas para as administrações que geraram os documentos. Assim, a construção do plano de classificação deve contemplar as relações entre os documentos, mas não só essas, também, as relações entre os documentos e as pessoas, assim, o sistema de classificação se tornaria uma ferramenta útil e facilitadora da descrição e avaliação dos documentos.

Assim, a classificação funciona como a base para todos os processos de organização de arquivos e evidentemente para a representação do conhecimento arquivístico.

3 A Descrição Arquivística: elementos históricos e conceituais

A palavra descrição é originada do termo latino *descriptio*, originado do termo *describere*, que significa transcrever, copiar, narrar, definir, distribuir, colocar em classes, escrever sobre. Portanto, o termo descrição arquivística significa, literalmente, **escrever sobre o material arquivístico** e abarca as ideias de representação, identificação e organização.

Essa atividade de descrever sobre os documentos de arquivo esteve presente desde os tempos mais remotos, e seus objetivos mudaram com o passar do tempo.

Na Antiguidade, por exemplo, repertórios de documentos eram construídos em tábuas de argila, e serviam para impedir a consulta direta aos documentos e manter o conteúdo dos arquivos em caso de deslocamento repentino ou em caso de guerra ou fogo. Neste caso, o ato de descrever estava ligado à preservação dos registros.

Na Idade Média, inventários eram compilados com o objetivo de fornecer provas da existência de documentos; ou armazenar o material preservado para os sucessores custodiadores. Aqui, além de preservar os documentos, a descrição tinha um objetivo jurídico (pois fornecia prova da existência dos documentos) e administrativo, uma vez que possibilitava ao custodiador dos documentos manter o controle sobre os fundos e facilitar a recuperação da informação para fins de valor primário).

> A necessidade de controle físico e administrativo foi sutilmente sentido, não apenas para efeito de uma recuperação efetiva e arranjo físico, mas também para garantir que os repositórios arquivísticos servissem sua função de " memória perpétua". Memória perpétua é um conceito jurídico segundo o qual os documentos preservados em arquivos são provas autênticas e permanentes de ações passadas. Essa ideia governou todo o trabalho arquivístico até o século XVIII, e foi a principal razão para a preservação, assim como para a descrição do material arquivístico (DURANTI, 1993, p. 49).

Apesar de o ato de descrever documentos exista desde a Antiguidade, o termo "descrição arquivística" e sua junção à classificação, enquanto duas atividades integradas, datam do século XIX, fruto do movimento historiográfico e científico da época que passam a entender o documento de arquivo como testemunho para os estudos do passado relacionado fundamentalmente com os documentos medievais.

Durante esse período, mais um objetivo é adicionado à atividade de descrever: fornecer acesso aos documentos. Com a abertura dos arquivos ao cidadão, depois da Revolução Francesa, percebe-se que mais pessoas além do criador e de alguns estudiosos acadêmicos, começam a usar o material arquivístico. Logo, faz-se necessário uma adaptação e redirecionamento da atividade de descrição, como destaca Duranti (1993, p. 50)

> [...] a descrição vem perdendo progressivamente seus objetivos de prestar contas dos fundos da instituição arquivística, facilitando a recuperação física de documentos e fornecendo acesso à informação. Esses objetivos são agora preenchidos pela compilação de registros de acesso, listas de localização, índices e todos os tipos de instrumentos de recuperação. O principal objetivo de escrever sobre os documentos é destacar as relações contextuais e de proveniência.

Portanto, a partir do século XX, a descrição começa a se preocupar menos com a criação de simples instrumentos de pesquisa e substitutos dos documentos, e passa a focar na relação do documento com seu contexto, trabalhando em paralelo com a classificação. Pode-se dizer, então, que descrever é um processo que permite ao arquivista representar – no sentido mais denotativo do termo - todo o conhecimento que é mantido em seu arquivo.

Para Rodrigues (2003, p. 217) a descrição é uma função-chave no que diz respeito à representação das informações arquivísticas e na possibilidade de acesso a elas. Esta representação vai além do documento, perpassando suas **funções** e seus **órgãos**

produtores. Neste sentido, o princípio de respeito aos fundos exerce, mais uma vez, um papel central no estabelecimento dos níveis que serão descritos.

A compreensão da descrição e classificação enquanto atividades conjuntas possibilita a existência de relações interdisciplinares entre a Arquivística e a Organização do Conhecimento, possibilitando à primeira novas abordagens teórico metodológicas.

4 Considerações

A Arquivística possui um desenvolvimento teórico recente no que se refere a suas bases conceituais e seu desenvolvimento metodológico. Buscou-se apresentar aspectos integradores da teoria Arquivística, passíveis de relações com a organização do conhecimento visando uma apropriação por parte de ambas as áreas, uma vez que, os arquivos tornam-se um espaço de aplicação de metodologias da organização do conhecimento e a Arquivística ganha um novo espaço para discussão de seus preceito histórico-conceituais.

Ambas as áreas só têm a ganhar com relações desta natureza e estudos como este devem correr com mais frequência e profundidade visando um aumento doa aparatos metodológicos neste universo.

Referências

DUCHEIN, M. Theoretical principles and practical problems of respect des fonds in Archival Science. **Archivaria**, 16 (summer), p. 64-82, 1983.

DURANTI, L. Origin and development of the concept of Archival Description. **Archivaria**, 35 (Spring), p. 47-54, 1993.

JENKINSON, H. **A manual of archive administration**: including the problems of war archives and archive making. Oxford: The Clarendon Press, 1922.
Hjørland, B. What is Knowledge Organization (KO)? **Knowledge Organization**. 35 (2) p. 86-102, 2008.

RODRIGUES, G. M. A representação da informação em arquivística: uma abordagem a partir da perspectiva da norma internacional de descrição arquivística. In: RODRIGUES, G.M.; LOPES, I.L.(Org.). **Organização e representação do conhecimento na perspectiva da ciência da informação**. Brasília: Thesaurus, 2003, v. 2, p. 210-230.

SABOURIN, P. Constructing a funcional-bases records classification system: business activity sctructure classification system. **Archivaria**, n. 51, p. 137-154, 2001.

SHEPHERD, E. & YEO, G. **Managing records**: a handbook of principles and practice. London: Facet, 2003.

Ontologias e Knowledge Organization System (SKOS): aproximações e diferenças

Rogério Aparecido Sá Ramalho
Universidade Federal de São Carlos
ramalho@ufscar.br

1 Introdução

Um dos principais desafios inerente à área de Organização do Conhecimento (OC) é o consenso sobre o que representar e como fazê-lo, de modo que diferentes áreas, (Ciência da Informação, Ciência da Computação, Linguística, Ciência Cognitivas, dentre outras), têm apresentado diversos modelos e instrumentos de representação, tornando a cada dia mais complexa sua distinção, tanto no campo teórico quanto prático.

No âmbito da área de Ciência da Informação para compreender a real importância das pesquisas em OC torna-se necessário considerar seu próprio significado e abrangência, a partir dos estudos realizados ao longo dos tempos que resultaram em um grande número de padrões e diretrizes, bem como em desenvolvimentos teóricos e avanços da tecnologia da informação (HJORLAND, 2003).

Ao longo dos últimos anos pesquisas relacionadadas ao desenvolvimento de ontologias têm favorecido a aproximação entre os instrumentos tradicionalmente utilizados no âmbito da Ciência da Informação com as novas tecnologias semânticas. “[...] as ontologias têm sido apontadas como uma nova possibilidade de representação de recursos informacionais em bibliotecas” (RAMALHO, 2010, p. 17). Neste contexto, a partir de agosto de 2009 com a publicação da Recomendação do *World Wide Web Consortium* (W3C) do *Simple Knowledge Organization System* (SKOS) este novo instrumento de representação tem chamado a atenção de pesquisadores, sendo, por vezes, comparado com ontologias,

fato que contribui para a geração de conflitos terminológicos e conceituais acerca das aproximações existentes entre tais instrumentos.

Partindo dessas premissas, a presente pesquisa realiza uma breve discussão de literatura, de caráter interdisciplinar, tendo como objetivo contribuir para um maior esclarecimento terminológico sobre as aproximações e diferenças existentes entre ontologias e SKOS, analisando as potencialidades de cada um destes instrumentos e seus reflexos no *corpus* teórico da área de Ciência da Informação.

2 Ontologias como instrumento de representação

No âmbito da área de Ciência da Informação o termo ontologia passou a ser utilizado no final da década de 1990, quando instrumentos de representação de informações despertaram maior interesse dos pesquisadores das mais diversas áreas do conhecimento, impulsionados pela criação do ambiente Web e desenvolvimento de novas tecnologias digitais (SOERGEL, 1999; VICKERY, 1997).

García-Marco (2007) ressalta que o crescente aumento do número de pesquisas relacionadas com ontologias e metadados na área de Ciência da Informação constitui a "ponta do *iceberg*" de um processo de reconfiguração disciplinar, como resultado da integração de diversas Ciências. Deste modo, verifica-se ao longo da última década um maior empenho na consolidação das definições de ontologia e análise das possibilidades de uso como instrumento de representação (SANTOS; CORREA; SILVEIRA, 2013).

Entre as definições de ontologia mais referenciadas na literatura destaca-se a apresentada por Gruber (1993, p. 1) de que "é uma especificação explícita de uma conceitualização". Contudo, tal definição remete o conceito de ontologia a aspectos filosóficos, favorecendo a geração de conflitos terminológicos, conforme destaca Guarino (1998, p. 5):

> No sentido filosófico, podemos nos referir a uma "ontologia" como um sistema particular de categorias que versa sobre uma certa visão do mundo. Desta forma, este sistema não depende de uma linguagem particular: a ontologia de Aristóteles é sempre a mesma, independente da linguagem usada para descrevê-la. Por outro lado, em seu uso mais prevalecente na IA [inteligência artificial], uma ontologia é referida como um artefato de

> engenharia, constituído de um vocabulário específico usado para descrever uma certa realidade e um conjunto de pressupostos explícitos relacionados com o significado pretendido para as palavras do vocabulário.

As ontologias figuram como uma nova abordagem computacional de representação, descrição e organização de conteúdos informacionais, favorecendo a aproximação dos modelos clássicos de representação, utilizados tradicionalmente no âmbito da área de Ciência da Informação, com as novas tecnologias semânticas desenvolvidas ao longo dos últimos anos. Quanto ao enfoque das representações semânticas, Zafalon (2013) esclarece que "é a semântica que dá conta dos processos mentais segundo os quais se produz, constitui, compreende e descreve a representação de um recurso informacional."

Ao longo dos últimos anos inúmeras metodologias têm sido propostas para o desenvolvimento de ontologias. Dentre elas, destacam-se a de Uschold e King (1995), a de Uschold e Gruninguer (1996), a de Férnandez, Gómez-Pérez e Juristo (1997), e o método 101, proposto por Noy e McGuinness (2001). No âmbito da área de Ciência da Informação, como substratos teóricos que podem ser utilizados para o desenvolvimento de ontologias, podemos destacar a teoria do conceito e a teoria da classificação facetada utilizadas na elaboração de tesauros (CAMPOS; GOMES, 2006; GOMES, et al., 1990), bem como a norma ANSI/NISO Z39.19-2005, que apresenta diretrizes para a elaboração de vocabulários controlados.

No intuito de aproximar o conceito de ontologia aos pressupostos da área de Ciência da Informação, principalmente quanto aos processos de representação, tratamento, organização e recuperação de recursos informacionais, recorre-se a Ramalho (2010, p. 107) que define ontologia como

> [...] um artefato tecnológico que possibilita representar formalmente as propriedades e relacionamentos de um determinado modelo conceitual, favorecendo a utilização de inferências automáticas nos processos de organização e recuperação de recursos informacionais.

Tal definição caracteriza as ontologias no âmbito da área de Ciência da Informação como um instrumento de nível epistemológico, concebido com o propósito de possibilitar a

realização de inferências automáticas, não sendo suficiente apenas representar os conceitos e relacionamentos existentes, mas sim garantir computabilidade das informações por meio da realização de inferências.

De modo a garantir tais caraterísticas, ontologias geralmente são implementadas a partir da *Web Ontology Language* (OWL) uma linguagem desenvolvida no bojo do projeto Web Semântica, visando garantir o máximo de expressividade sem perda de completude computacional e que pudesse ser inserida em páginas Web. OWL permite representar computacionalmente classes, propriedades, indivíduos e valores de dados, que são armazenados como documentos Web.

Em sua primeira versão a Linguagem OWL teve como propósito principal permitir a representação de classes, indivíduos e relações existentes. Mais recentemente, com a publicação da OWL 2, novas funcionalidades foram incorporadas, possibilitando expressar restrições adicionais sobre propriedades, novas características de propriedades, incompatibilidade de propriedades, cadeias de propriedades e propriedades-chave (W3C Recommendation, 2012).

3 Simple Knowledge Organization System (SKOS)

SKOS é um modelo de dados que possibilita expressar Sistemas de Organização do Conhecimento (SOC), tais como tesauros, sistemas de classificação, e taxonomias em um formato computacional “legível” por máquinas. O SKOS foi desenvolvido a partir de sucessivos projetos e grupos de trabalho do *World Wide Web Consortium* (W3C) e adotado em 2009 como Recomendação do W3C para dar suporte a representação, uso e interoperabilidade entre diferentes tipos de vocabulários (ZOGHLAMI; KERHERVÉ; GERBÉ, 2011).

A partir da utilização de SKOS, conceitos podem ser identificados utilizando URIs, rotulados com sequencias textuais em uma ou mais línguas naturais, notações embutidas, documentadas com vários tipos de nota, ligados a outros conceitos organizados em hierarquias informais e redes de associação, agregados em esquemas conceituais, agrupados em coleções rotuladas e/ou ordenadas ou mapeados em conceitos de outros esquemas (W3C Candidate Recommendation, 2009).

Para a formalização dos dados representados SKOS é utilizado o padrão *Resource Description Framework* (RDF). Apesar de possuir uma expressividade limitada, RDF possibilita a representação formal de declarações sobre objetos, fornecendo mecanismos para a descrição de relacionamentos de recursos específicos, limitando-se ao nível de instância e não permitindo a utilização de conectivos lógicos (RAMALHO; OUCHI, 2011).

Como uma aplicação do RDF, SKOS permite que conceitos possam ser representados na Web, interligando dados e integrando outros esquemas conceituais. Cada conceito SKOS é definido como um recurso RDF e está ligaco com outros conceitos e propriedades . Assim, pode-se afirmar que o modelo SKOS é baseado em uma visão de vocabulário, onde os objetos primários não são rótulos, em vez disso, são conceitos representados por rótulos, os quais podem ser relacionados com outros padrões de metadados e esquemas de representação.

4 Considerações Finais

Observa-se que a construção de ontologias busca a formalização relações entre conceitos que possam ser processados computacionalmente, de modo que estruturas de dados heterogêneas possam ser “compreensíveis” computacionalmente, favorecendo “interpretações” lógicas a partir de inferências automáticas.

Assim, o desenvolvimento de ontologias exige um maior envolvimento de comunidades de especialistas, de modo que as informações expressas na ontologia possam representar computacionalmente um consenso de uma comunidades específica.

Em relação ao modelo de dados SKOS, verifica-se que o mesmo caracteriza-se como uma maneira inovadora de representar computacionalmente instrumentos e esquemas de representação tradicionalmente utilizados na área de Ciência da Informação, como tesauros, sistemas de classificação, e taxonomias. Contudo, devido aos seu propósito principal, de estabelecer um vocabulário simples que favoreça o mapeamento de conceitos existes em diferentes esquemas conceituais, a representatividade computacional do modelo SKOS é limitada.

Deste modo, observa-se que propósito de SKOS é bastante diferente da expressividade almejada pelas ontologias, não devendo tais instrumentos serem

confundidos. Contudo, apesar de suas limitações o modelo SKOS destaca-se devido às potencialidades oferecidas para a intereperabilidade de recursos informacionais. Espera-se que os mapeamentos conceituais a partir do padrão SKOS sejam uma das principais contribuições para o futuro da Web Semântica, interligando de forma prática diferentes esquemas conceituais.

Assim, observa-se que as ontologias ainda figuram como o mais expressivo intrumento de representação, capaz de engendrar uma nova "constelação" epistemológica no âmbito da Organização do Conhecimento, apontando perspectivas inovadoras e novas possibilidades no fazer profissional da área de Ciência da Informação na medida que exigem maior rigor metodológico no estabelecimento de suas relações formais, a fim de favorecer a realização de inferências automaticas.

REFERÊNCIAS

FÉRNANDEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. **Methontology**: From Ontological Art Towards Ontological Engineering. Palo Alto: Association for the Advancement of Artificial Intelligence, 1997. AAAI Technical Report SS-97-06. Disponível em: <http://oa.upm.es/5484/1/METHONTOLOGY_.pdf>. Acesso em: 20 jun. 2014.

HJORLAND, B. Fundamentals of knowledge organization. Knowledge Organization, v.30, n.2, p. 87-111, 2003.

GARCÍA MARCO, F. J. Ontologías y organización del conocimiento: retos y oportunidades para el profesional de la información. **El profesional de la información**, v. 16, n. 6, p. 541- 550, Nov.-Dic. 2007. Disponível em: <http://www.elprofesionaldelainformacion.com/contenidos/2007/noviembre/01.pdf>. Acesso em: 10 jan. 2014.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2007.

GOMES, H. E. et al. **Manual de elaboração de tesauros monolíngues**. Brasília: CNPq/PNBU, 1990.

GRUBER, T. R. Toward principles for the design of ontologies used for knowledge sharing. In: GUARINO, N.; POLI, R. (Org.). **Formal Ontology in Conceptual Analysis and Knowledge Representation**. [s.l.]: Kluwer Academic Publishers, 1993. Disponível em: <http://pdf.aminer.org/000/912/413/toward_principles_for_the_design_of_ontologies_used_for_knowledge.pdf>. Acesso em: 29 jun. 2014.

GUARINO, N. **Formal Ontology and Information Systems**. [Vilnius]: Vilnius University Faculty of Mathematics and Informatics, 1998. Disponível em: <http://uosis.mif.vu.lt/~donatas/Vadovavimas/Temos/OntologiskaiTeisingasKoncepcinisMod

eliavimas/papildoma/Guarino98-Formal%20Ontology%20and%20Information%20Systems.pdf>. Acesso em: 26 jul. 2014. Amended version of a paper appeared in N. Guarino (ed.), Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, pp. 3-15.

NOY, N. F.; MCGUINESS, D. L. **Ontology development 101**: a guide to creating your first ontology. Stanford: Standford University, [2001?]. Disponível em: <http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy- mcguinness.html>. Acesso em: 20 fev. 2014.

RAMALHO, R. A. S. **Desenvolvimento e utilização de ontologias em bibliotecas digitais**: uma proposta de aplicação. 2010. 145f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2012.

RAMALHO, R. A. S.; FUJITA, M. S. L. Aplicabilidad de ontologías en bibliotecas digitales. **Anales de Documentación**, v. 14, n. 1, 2011. Disponível em: <http://revistas.um.es/analesdoc/article/view/120161>. Acesso em: 12 mar. 2014.

RANGANATHAN, S. R. **As cinco leis da biblioteconomia**. Brasília: Briquet de Lemos, 2009.

SANTOS, M. T.; CORREA, R. F.; SILVEIRA, M. A. A. Estudos brasileiros sobre ontologia na Ciência da Informação. **Datagramazero**, v. 14, n. 1, 2013. Disponível em: <http://www.dgz.org.br/fev13/Art_05.htm>. Acesso em: 20 jun. 2014.

SILVEIRA, D. T.; CÓRDOVA, F. P. A pesquisa científica. In: GERHARDT, T. E.; SILVEIRA, D. T. (Orgs.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

SOERGEL, D. The rise of ontologies or the reinvention of classification. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1119-1120, Oct. 1999. Disponível em: <http://www.dsoergel.com/cv/B70.pdf>. Acesso em: 10 jan. 2014.

SVENONIUS, E. **The Intellectual Foundation of Information Organization**. Cambridge: MIT Press, 2000.

TRIVIÑOS, A. N. S. **Introdução à pesquisa em ciências sociais**: a pesquisa qualitativa em educação. São Paulo: Atlas, 1987.

USCHOLD, M.; GRUNINGER M. Ontologies: principles, methods and applications. **The Knowledge Engineering Review**, v. 11, n. 2, p. 93-136, 1996.

USCHOLD, M.; KING, M. Towards a Methodology for Building Ontologies. In: IJCAI-95: Workshop on Basic Ontological Issues in Knowledge Sharing, 1995, Montreal, Canada. **Proceedings**. 1995. Disponível em: <http://www1.cs.unicam.it/insegnamenti/reti_2008/Readings/Uschold95.pdf>. Acesso em: 20 jun. 2014.

VICKERY, B. C. Ontologies. **Journal of Information Science**, v. 23, n. 4, p. 277–286, 1997.

W3C Recomentadion. **OWL 2 Web Ontology Language Structural Specification and Functional-Style Syntax** (Second Edition). 2012. Motik, B, Patel-Schneider, P.F. and Parsia, B. eds. Disponível em: <http://www.w3.org/TR/owl2-syntax/>. Acesso em: 20 mar. 2015.

W3C Candidate Recomentadion. SKOS **Simple Knowledge Organization System Reference**. 2009. Disponível em: <http://www.w3.org/TR/2009/CR-skos-reference-20090317/>. Acesso em: Março, 2015.

ZAFALON, Z. R. Simbiose entre catalogação e recursos tecnológicos: associações entre Ciência da Informação e Linguística pelo viés dos registros bibliográficos automatizados. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. [**Anais eletrônicos**]. 2013. Disponível em: <http://enancib2013.ufsc.br/index.php/enancib2013/XIVenancib/paper/download/496/328>. Acesso em: 20 maio 2014.

ZOGHLAMI, K.; KERHERVÉ, B.; GERBÉ, O. Using a SKOS engine to create, share and transfer terminology data sets. In: INTERNATIONAL CONFERENCE ON SIGNAL IMAGE TECHNOLOGY & INTERNET-BASED SYSTEMS, 7th, 2011, Dijon, France. **Proceedings**… Los Alamitos, California: IEEE Computing Society, 2011. p.46-53.

ZHAO, Y; HU, C. An Ontology-based Framework for Knowledge Service in Digital Library. In: INTERNATIONAL CONFERENCE ON WIRELESS COMMUNICATIONS, NETWORKING AND MOBILE COMPUTING, 2007, Shangai. **Proceedings**. Palo Alto: IEEE, 2007. p. 5345-5348.

Estudos de genealogia acadêmica como abordagem para análise de domínio

Renata Cristina Gutierres Castanha
Universidade Estadual Paulista
regutierres@gmail.com

Maria Cláudia Cabrini Grácio
Universidade Estadual Paulista
cabrini@marilia.unesp.br

1 Introdução

No âmbito da Ciência da Informação, define-se domínio como uma comunidade discursiva ou de pensamento que participa da partilha das atividades, na qual a linguagem, a estrutura de trabalho e padrões de cooperação, as formas de comunicação e organização do conhecimento, os sistemas de informação e os critérios de relevância, são reflexos dos objetos de trabalho dessa comunidade e do seu papel na sociedade (HJØRLAND; ALBRECHTSEN, 1995).

A Análise de Domínio constitui relevante abordagem para caracterização da ciência que permite identificar e analisar as condições pelas quais o conhecimento científico se constrói e se socializa, na medida em que propicia observar a integração do indivíduo com o contexto social das comunidades em que está inserido e os conceitos de informação e conhecimento têm seu significado segundo a compreensão compartilhada dos membros dessas comunidades (OLIVEIRA; GRÁCIO, 2013).

Nesse contexto, Hjørland (2002) sugere 11 abordagens para se estudar e conhecer um domínio e destaca que a utilização conjunta dessas abordagens proporciona um entendimento mais profundo do domínio a ser estudado. Entre as abordagens apresentadas, segundo o autor, os estudos bibliométricos operacionalizam e sustentam as análises qualitativas, que buscam compreender as dimensões sócio-

cognitivas da ciência, propiciando análises consistentes do modo como o conhecimento é produzido, organizado e disseminado.

Desse modo, os estudos quali-quantitativos da construção dos espaços intelectuais científicos ocupam lugar fundamental na dimensão investigativa dos diversos domínios do conhecimento, tornando possível verificar o que é efetivamente importante ou significativo em um dado campo, de tal modo que aspectos como tendências, padrões, processos, agentes e seus relacionamentos possam ser identificados e analisados (DANUELLO, 2007).

Hjørland (2002) propôs a abordagem bibliométrica focando, principalmente, nos estudos de citação. Todavia, vários estudiosos argumentam que, atualmente, as análises quantitativas de um domínio vão além dos estudos bibliométricos, em decorrência, principalmente, do surgimento de novas tecnologias para a recuperação, organização e análise de grandes quantidades de dados (SUGIMOTO, 2014; HERRERO-SOLANA, 2001), fazendo surgir novos procedimentos de pesquisa e análise, entre eles as Análises de Redes Sociais, e mais recentemente, em particular, as análises de Genealogia Acadêmica, que contribuem para a apreensão de aspectos cognitivos da informação.

A análise da Genealogia Acadêmica é constituída de estudos da herança intelectual, operacionalizada pelas relações interdependentes entre os alunos e seus orientadores (SUGIMOTO, 2014). Consiste em um tipo específico de rede social, caracterizada como árvore, e de colaboração científica, uma vez que seus atores são os orientadores e orientandos.

Considerando o acima exposto, neste estudo objetiva-se, de forma geral, descrever e compreender a abordagem da análise da Genealogia Acadêmica para os estudos de Análise de Domínio.

De forma mais específica, propõe-se discutir os aspectos fundamentais da análise da Genealogia Acadêmica e a possível complementaridade e articulação com as demais abordagens da teoria da Análise de Domínio proposta por Hjørland (2002). Ainda, busca apresentar aplicações da análise da Genealogia Acadêmica na literatura científica, como abordagem para caracterização de um domínio científico.

Este estudo apresenta caráter teórico e busca contribuir para a melhor compreensão conceitual-metodológica da fundamentação da análise da Genealogia

Acadêmica. Ainda, apresenta natureza documental ao trazer exemplos de aplicações dessa abordagem na literatura científica, para a identificação e caracterização de um domínio.

2 Genealogia acadêmica como abordagem para a análise de domínio

A necessidade de compreender as estruturas sociais dentro da ciência trouxe metodologias analíticas como a Análise de Redes Sociais, que conta com contribuições da sociologia da ciência, da matemática e da história social do conhecimento, a fim de investigar o papel e a relação social entre os atores em um domínio (BUFREM; GABRIEL JÚNIOR; SORRIBAS, 2011).

A Genealogia Acadêmica, por meio do seu esquema de rede social em formato de árvore, identifica, evidencia e organiza os cientistas e acadêmicos, segundo suas relações de orientação ou supervisão concluídas, contribuindo para a visualização de uma linhagem científica (MIYAHARA, 2011). Expõe um tipo especial de colaboração científica em um domínio, em função da contribuição especializada do orientador para alcançar os objetivos da pesquisa do orientando, pela qual o orientando adquire novas habilidades e conhecimento tácito nessa atividade colaborativa (BALANCIERI et al., 2004).

As ideias, contribuições e pontos de vista de um cientista podem ser perpetuados e continuar a influenciar o pensamento científico nas gerações seguintes de pesquisadores, por meio dos acadêmicos que orienta (ANDRAOS, 2005).

Uma árvore de Genealogia Acadêmica pode ser descendente e ascendente. Uma árvore descendente exibe, a partir de um pesquisador, seus orientandos e os eventuais orientandos dos orientandos. Por outro lado, uma árvore de genealogia ascendente, a partir de um pesquisador, exibe o seu orientador e possivelmente o orientador do orientador, seguindo assim, sentido contrário da árvore descendente (MIYAHARA, 2011).

Pode-se dizer, assim, que um estudo genealógico acadêmico descendente consiste em uma análise diacrônica de um domínio científico e o estudo genealógico ascendente em uma análise sincrônica de um domínio.

A análise genealógica acadêmica, forma particular de rede social, firma-se em um paradigma social, uma vez que a atividade científica apresenta um modelo social hierárquico, a partir das relações entre os pesquisadores e da forma que se consolida a transmissão de conhecimentos e práticas epistêmicas por meio dessas relações. Especificamente, no vínculo orientador-orientando, observa-se uma relação hierárquica social de interdependência, em que há uma forma particular de disseminação e transferência de conhecimento, que se reflete na comunidade discursiva e carrega a subjetividade existente no domínio em que estão inseridos.

Entende-se que essas práticas de orientação são responsáveis pela transferência tácita e explicita de conhecimento durante as interações interdependentes entre o orientador e o orientando (GIRVES; WEMMERUS, 1988).

Nesse sentido, a Genealogia Acadêmica caracteriza-se como uma abordagem fundamental para estudar a transferência de conhecimento em um domínio, fornecendo um meio para medir e analisar estas interações interdependentes na esfera acadêmica. Embora a vida do acadêmico seja finita, suas contribuições científicas são impactadas por meio de sucessivas gerações de orientandos (SUGIMOTO, 2014).

Desse modo, as árvores e florestas de Genealogia Acadêmica, por meio das relações de orientação, contribuem para a visualização e caracterização dos grupos e correntes científicas atuantes em um domínio.

Considera-se, nesse contexto, que os estudos genealógicos associados aos estudos históricos e epistemológicos, propostos por Hjørland (2002), permitem uma visualização ampla, objetiva e contextualizada de um domínio científico.

Observa-se, ainda, que os estudos de genealogia acadêmica se revelam significativos para a avaliação do impacto das orientações acadêmicas no desenvolvimento científico de uma comunidade e na identificação dos principais atores (ROSSI; MENA-CHALCO, 2014).

Nesse contexto, esses estudos articulados aos estudos bibliométricos, propostos por Hjørland (2002), oferecem uma importante contribuição para a Análise de Domínio ao permitir conhecer e caracterizar tendências, padrões, processos, pensamentos dominantes, atores importantes e seus relacionamentos.

A fim de introduzir os estudos de Genealogia Acadêmica como abordagem para a Análise de Domínio, elaborou-se um diagrama (Figura 1) das 11 abordagens propostas por Hjørland (2002), acrescido dos estudos de Genealogia Acadêmica, com as associações propostas pelo autor, representadas pelos vetores, acrescidas das associações consideradas para os estudos genealógicos.

Primeiramente, observa-se que as abordagens histórica e epistemológica e crítica, embora sejam indispensáveis para a maioria das outras abordagens, não necessitam, segundo Hjørland (2002) de abordagens complementares, uma vez que o autor considera que a abordagem histórica seja um método substancial para o entendimento de um domínio, por oferecer uma perspectiva profunda e coesa, e a abordagem epistemológica e crítica, para a análise de um domínio, oferece conhecimento sobre as bases teóricas e a evolução da área a ser estudada.

Ainda, Hjørland (2002) não associa a abordagem sobre cognição científica, conhecimento especializado e inteligência artificial a nenhum outro tipo de abordagem proposta.

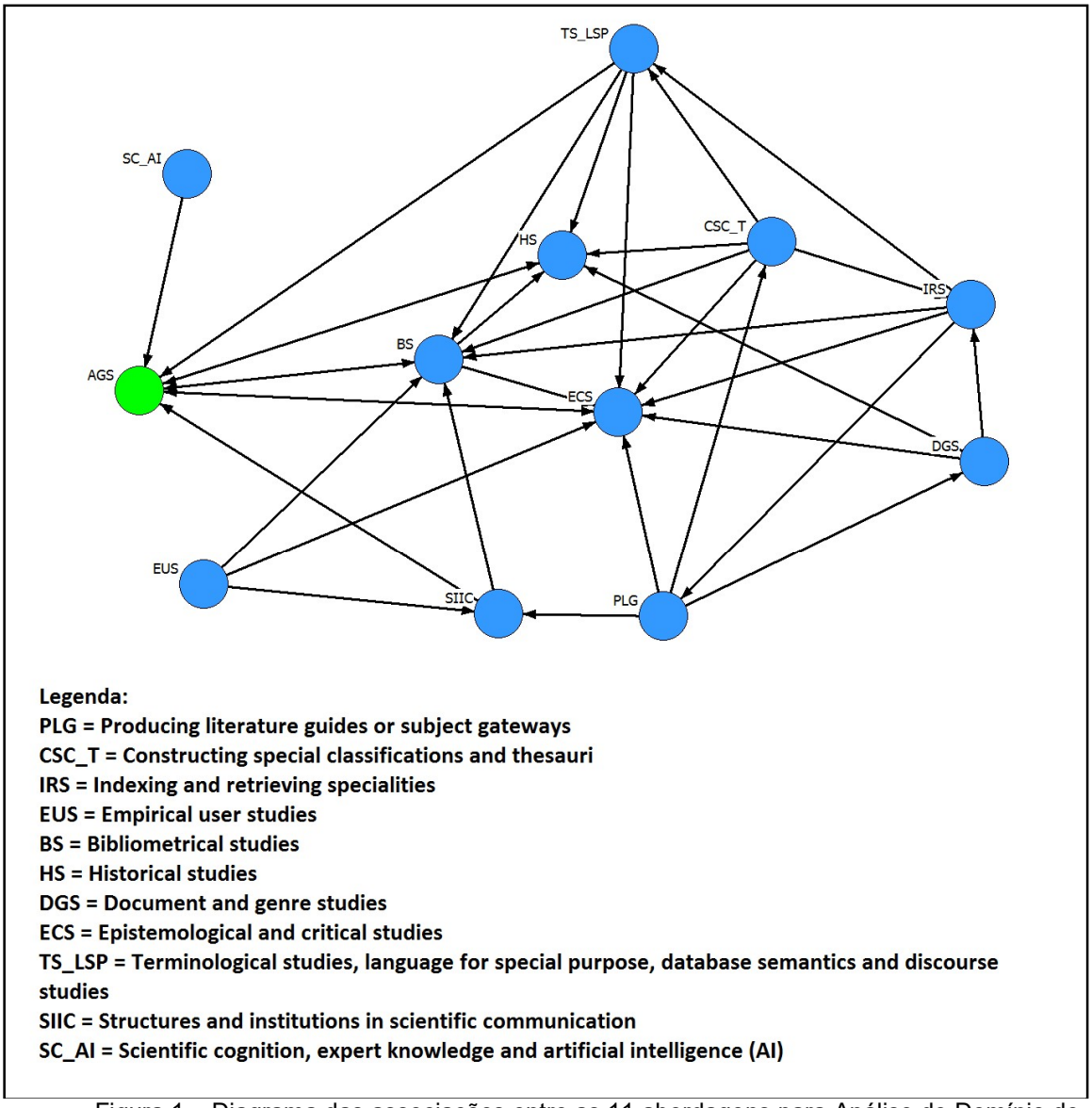


Figura 1 – Diagrama das associações entre as 11 abordagens para Análise de Domínio de Hjørland (2002), acrescido dos estudos de Genealogia Acadêmica (AGS).

Fonte: Elaboração das autoras.

Em relação à abordagem dos estudos de Genealogia Acadêmica, considera-se que sua análise pode ser enriquecida quando associada aos estudos históricos, bibliométricos e epistemológicos e críticos, contribuindo para a consolidação do entendimento significativo do domínio estudado.

Por outro lado, considera-se também que os estudos bibliométricos, os estudos históricos, estudos epistemológicos e críticos, os estudos de terminologia, os estudos das

estruturas e instituições e ainda, os estudos de cognição e de conhecimento especializado, podem se beneficiar quando associados aos estudos de Genealogia Acadêmica, enriquecendo o entendimento das origens, evolução e disseminação do conhecimento no domínio e contribuindo para a compreensão da formação de correntes de pensamento e tendências.

Diversos estudos analisando diferentes domínios científicos têm utilizado a abordagem da análise de Genealogia Acadêmica, a saber: Matemática (CHANG, 2011; MALMGREN et al., 2010; ROSSI; MENA-CHALCO, 2014), Ciência da Informação (MARCHIONINI et al., 2007; RUSSELL; SUGIMOTO, 2009; SUGIMOTO et al., 2011), Neurociência (DAVID; HAYDEN, 2012), entre outros. Observa-se que são estudos muito recentes e, em sua maioria, realizados a fim de analisar um domínio, identificando como a comunidade discursiva reconhece a sua referência teórica. Destaque-se, todavia, que, nesses estudos, as bases teóricas de análise de domínio não foram explicadas, ou seja, estes pesquisadores, mesmo realizando uma análise de domínio, não explicitam ou têm ciência de que a estão realizando.

Salienta-se, ainda, que muito do que tem sido publicado, na literatura científica, sobre Genealogia Acadêmica, tem mostrado que suas características são baseadas em grafos e redes sociais, uma vez que a Genealogia Acadêmica constitui um tipo especial de rede, hierárquica e definida como árvore (genealógica).

3 Considerações finais

Este estudo objetivou contribuir para a melhor compreensão conceitual-metodológica dos estudos de genealogia acadêmica, que envolvem a análise da relação hierárquica interdependente entre orientador e orientando, e evidenciar a contribuição desses estudos como abordagem para Análise de Domínio, os quais, em cooperação com outras abordagens propostas por Hjørland (2002), possibilitam conhecer e caracterizar um domínio científico.

Considera-se que, ao evidenciar as origens, evolução e disseminação do conhecimento em um domínio científico, por meio das redes de relações de supervisão e orientação, e contribuir para a compreensão da formação de grupos, correntes de

pensamento e tendências nesse domínio, os estudos de Genealogia Acadêmica constituem relevante abordagem de Análise de Domínio.

Nesse sentido, pondera-se que os estudos de Genealogia Acadêmica constituem também um importante aporte para os estudos da sociologia da ciência, ao permitir uma visualização sincrônica e diacrônica das relações entre pesquisadores e acadêmicos de um domínio científico.

Finalizando, considera-se que análise da Genealogia Acadêmica pode consistir em uma abordagem analítica significativa para os estudos dos domínios científicos, articulada às análises quantitativas, como os estudos bibliométricos e às análises qualitativas, como estudos históricos, epistemológicos e críticos, para uma melhor caracterização e compreensão de um determinado domínio.

Referências

ANDRAOS, J. Scientific genealogies of physical and mechanistic organic chemists. *Canadian Journal of Chemistry*, v.*83*, n.9, p.1400–1414, 2005.

BALANCIERI, R.; BOVO, A.; KERN, V.; PACHECO, R. C.; BARCIA, R. M. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. *Ciência da Informação*, v.34, n.1, 2005.

BUFREM, L., GABRIEL JÚNIOR, R., SORRIBAS, T. Redes sociais na pesquisa científica da área de ciência da informação. *DataGramaZero*, v.12, n.3, 2011.

CHANG, S. *Academic genealogy of mathematics*. Singapore: World Scientific Publishing, 2011.

DANUELLO, J. *Produção científica docente em tratamento temático da informação no Brasil: uma abordagem métrica como subsídio para a análise do domínio*. 122 p. Dissertação (mestrado) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2007.

DAVID, S.; HAYDEN, B. Neurotree: A collaborative, graphical database of the academic genealogy of neuroscience. *PLoS ONE*, v.*7*, n.10, 2012.

GIRVES, J.; WEMMERUS, V. Developing models of graduate student degree progress. *Journal of Higher Education* , v.59, n.2, 163–189, 1988.

HERRERO-SOLANA, V. Producción científica de la Universidad Nacional de Mar del Plata: análisis de dominio. *Nexos*, v.8, n.14, p. 4–10, 2001.

HJØRLAND, B. Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, v.58, n.4, p.422–462, 2002.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in information science: Domain-analysis. *Journal of American Society Information Science*, v.46, p.400–425, 1995.

MALMGREN, R.; OTTINO, J.; AMARAL, L. A. The role of mentorship in protégé performance. *Nature*, v.465, n.3, p.622–627, 2010.

MARCHIONINI, G.; SOLOMON, P.; DAVIS, C.; RUSSELL, T. Information and library Science MPACT: a preliminary analysis. *Library & Information Science Research*, v.28, n.4, p.480–500, 2006.

MIYAHARA, E.K. *Genealogia Acadêmica Lattes***.** Monografia (Bacharelado em Ciência da Computação) – Universidade de São Paulo, Instituto de Matemática e Estatística, 2011.

OLIVEIRA, E. T; GRACIO, M. C. C. Studies of Author Cocitation Analysis: a bibliometric approach for domain analysis. *Iris*, v.2, n.1, p.12-23, 2013.

ROSSI, L.; MENA-CHALCO, J. Caracterização de árvores de genealogia acadêmica por meio de métricas em grafos. In: *Brazilian Workshop on Social Network Analysis and Mining,* Brasília-Brazil, p.1–12, 2014.

RUSSELL , T.; SUGIMOTO, C. R. MPACT family trees: Quantifying academic genealogy in library and information science. *Journal of Education for Library and Information Science*, v.50, n.4, p.248–262, 2009.

SMIRAGLIA, R. ISKO 11's Diverse Bookshelf: an editorial. *Knowledge Organization*, v. 38, n.3, p.179–186, 2011.

SUGIMOTO, C.R. Academic Genealogy. In: CRONIN, B.; SUGIMOTO, C.R.(Eds.). *Beyond bibliometrics: harnessing multidimensional indicators at scholarly impact.* Cambridge: MIT Press, 2014.

SUGIMOTO, C. R.; NI , C.; RUSSELL , T. G.; BYCHOWSKI, B. Academic genealogy as an indicator of interdisciplinarity: An examination of dissertation networks in library and information science. *Journal of the American Society for Information Science and Technology*, v.62, n.9, p.1808–1828, 2011.

A dimensão epistemológica da análise documental de conteúdo de obras de ficção na organização do conhecimento

Roberta Caroline Vesú Alves
Universidade Estadual Paulista
robertavesu@gmail.com

João Batista Ernesto de Moraes
Universidade Estadual Paulista
jota@marilia.unesp.br

1 Introdução

A interdisciplinaridade da Análise Documental de Conteúdo para textos narrativos ficcionais engloba alguns aspectos teóricos das áreas da Análise do Discurso e Semiótica de Greimas, que subsidiam seus procedimentos de análise, síntese e representação (FIORIN, 2008; GREIMAS, 1971; GUIMARÃES; MORAES; GUARIDO, 2007).

Esses aspectos epistemológicos incorporados à Análise Documental de Conteúdo colaboram para seu embasamento teórico-metodológico e desenvolve áreas em que mantém interlocução teórica, como o Tratamento Temático da Informação, Linguística Documental, Textologia Documental e Organização do Conhecimento.

Diante disso, o objetivo deste trabalho consistiu em analisar a interdisciplinaridade da Análise Documental de Conteúdo de textos narrativos ficcionais, estudada por pesquisadores da Universidade Estadual Paulista (UNESP) de Marília-SP-Brasil, verificando seus aspectos epistemológicos, também relacionados à Organização do Conhecimento. Para isso, utilizou-se metodologia de estudo exploratório interdisciplinar, visando obter análise qualitativa das informações.

Verificou-se que a Análise Documental de Conteúdo para essas obras apresenta aspectos interdisciplinares que subsidiam seus procedimentos, caracterizados

basicamente como semiótico greimasiano e linguísticos no âmbito do texto e discurso, que enriquecem epistemologicamente também outras áreas, conforme a seguir.

2 Análise Documental de Conteúdo e a organização do conhecimento

De acordo com Gardin et al (1981), a Análise Documental de Conteúdo se preocupa com procedimentos de representação dos documentos de forma distinta do original, por meio de classificação, tradução, resumo e indexação, com a função de facilitar a recuperação.

Para Guimarães, Moraes, Guarido (2007), a Análise Documental de Conteúdo desenvolve os procedimentos de análise, síntese e representação dos documentos para fins de recuperação pelo usuário, mediante as etapas analítica (procedimentos de leitura de partes importantes do documento, com estratégias metacognitivas para identificação de conceitos) e sintética (seleção de conceitos, condensação em resumo e representação em índices).

Além disso, conforme Guimarães (2008, 2009), consiste em corrente teórica do Tratamento Temático da Informação, de origem francesa, teve início no final da década de 1960 com estudos de Gardin e Coyaud, também é denominada de Análise Documental (*analyse documentaire*). Segundo Guimarães et al (2012), a Análise Documental de Conteúdo na França e no Brasil designa o Tratamento Temático da Informação, enquanto que na Espanha se insere no Tratamento Temático (descrição de forma e conteúdo).

Guimarães et al (2012, p.197), afirma que a Análise Documental de Conteúdo

> [...] encontrou solo fértil para desenvolvimento notadamente no Brasil e na Espanha, a partir dos anos 80: no Brasil, por meio estudos teóricos do Grupo TEMMA, de forte influência gardiniana, liderado por Johanna Smit, e na Espanha, a partir de uma forte conexão de autores como López Yepes e Sagredo Fernández, em Madrid, com a tradição otletiana ou ainda Nuria Amat, em Barcelona, com as idéias de Chaumier.

O Tratamento Temático da Informação, segundo Guimarães (2008), contém mais duas correntes teóricas em sua base epistemológica, a catalogação de assunto (*subject*

cataloguing) norte-americana e a indexação (*indexing*) de origem inglesa, que contribuem para sua consolidação teórica e metodológica.

Por exemplo, nos Estados Unidos e Inglaterra os estudos foram voltados para produtos e instrumentos de informação, e na França houve a "[...] preocupação com o desenvolvimento de referenciais teórico-metodológicos para o processo de Tratamento Temático da Informação propriamente dito", pois, os procedimentos até então utilizados eram guiados apenas pelo bom senso (GUIMARÃES, 2009, p.109).

Nesse contexto, o Tratamento Temático da Informação estuda a descrição e representação de forma e conteúdo dos documentos para recuperação, se insere no Tratamento da Informação, uma subárea da Biblioteconomia e Documentação que mantém interlocução teórica com a Organização do Conhecimento (BARITÉ, 2001).

Segundo Barité (2001), a Organização do Conhecimento está preocupada com o desenvolvimento de técnicas de construção, gestão, uso e avaliação para linguagens documentais, mediante premissas que determinam o conhecimento estabelecido socialmente a partir da informação e transformado em informação quando socializado, necessitando de organização para diferentes usos, entre outros aspectos (BARITÉ, 2001).

Nesse sentido, Esteban Navarro e García Marco (1995, p.149, tradução nossa), afirmam que o campo da Organização do Conhecimento consiste em,

> [...] disciplina dedicada ao estudo e desenvolvimento dos fundamentos e técnicas de planejamento, construção, gestão, uso e avaliação de sistemas de descrição, catalogação, ordenação, classificação, armazenamento, comunicação e recuperação dos documentos criados pelo homem para testemunhar, conservar e transmitir seu saber e seus atos, a partir de seu conteúdo, com o fim de garantir sua conversão em informação capaz de gerar novo conhecimento. Trata, portanto, de una ciência tridimensional, que se ocupa dos princípios, métodos e instrumentos postos em ação para a gestão do conhecimento humano desde uma tripla perspectiva: sua representação, sua organização e sua comunicação documental.

Diante disso, verifica-se que a Organização do Conhecimento se preocupa com aspectos de organização e representação necessários para socialização dos conhecimentos também materializados em documentos, entre outros aspectos. Contudo,

de modo subjacente à representação estão os procedimentos de análise e síntese dos conteúdos de documentos, o que envolve estudos de todas as etapas da Análise Documental de Conteúdo na Organização do Conhecimento.

3 Aspectos epistemológicos interdisciplinares da Análise Documental de Conteúdo para textos narrativos ficcionais

Os aspectos teórico-metodológicos interdisciplinares de procedimentos da Análise Documental de Conteúdo, para Guimarães (2009), abrangem áreas como a Linguística, Terminologia, Lógica, Psicologia Cognitiva, entre outras. Nesse contexto, foram verificados os aspectos teórico-metodológicos interdisciplinares que subsidiam epistemologicamente a Análise Documental de Conteúdo de textos narrativos ficcionais em estudos de pesquisadores da UNESP de Marília-SP-Brasil.

Esses estudos utilizam a aplicação do Percurso Gerativo de Sentido para identificação do *aboutness* desses textos, conforme Moraes (2011, 2012a, 2012b), Moraes e Guimarães (2006), Guimarães, Moraes e Guarido (2007), entre outros. Portanto, utilizam o Percurso Gerativo de Sentido como estratégia metacognitiva de identificação das partes importantes do texto para análise do conteúdo temático.

O Percurso Gerativo de Sentido consiste no modelo teórico para entendimento de diferentes níveis estruturais internos dos textos verbais e não-verbais e suas camadas de significação (aspectos sintáxicos e semânticos), como os narrativos e científicos, devido à narratividade inerente. Também, pertence à área da Análise do Discurso e Semiótica Greimasiana, tem como base os postulados de Greimas e estudos de Propp, Hjelmslev, entre outros, portanto, com características estruturalistas devido às interrelações entre as estruturas abstratas do texto e seus significados (GREIMAS, 1971; FIORIN, 2008).

O referido Percurso Gerativo apresenta as estruturas profunda ou fundamental, narrativa (manipulação, competência, performance e sanção) e discursiva (temas e figuras), e os aspectos semânticos entendidos por isotopia (FIORIN, 2008). Por isso,

permite a identificação do *aboutness* ou tema principal, que se apresenta em estrutura profunda (MORAES, 2011, 2012a, 2012b).

Verificou-se, portanto, que esse modelo teórico permitiu a identificação do *aboutness* proveniente das regras do texto em Análise Documental de Conteúdo (HUTCHINS, 1977; BEGHTOL, 1986; MORAES, 2011, 2012a, 2012b). Também, proporcionou a classificação dos temas do texto narrativo ficcional para finalidades de construção de índices e indexação de conceitos (GUIMARÃES; MORAES; GUARIDO, 2007).

Além disso, o Percurso Gerativo de Sentido atua na Análise Documental de Conteúdo, assim como a Linguística Textual de Van Dijk (2000) estabelece os parâmetros de estrutura e semântica do texto, por meio das microestruturas, macroestruturas, superestruturas textuais e macrorregras. Portanto, demonstrando aspectos textuais importantes para analise, síntese e representação.

A interdisciplinaridade da Análise Documental de Conteúdo para textos narrativos ficcionais tem como base o modelo teórico do Percurso Gerativo de Sentido, o que a caracteriza teoricamente e metodologicamente com aspectos semióticos greimasianos e linguísticos no âmbito do texto e discurso. Portanto, também amplia as perspectivas teóricas e epistemológicas do Tratamento Temático da Informação e do campo da Organização do Conhecimento, por meio da interdisciplinaridade com essas áreas.

Vale destacar também que os estudos em Análise Documental de Conteúdo voltados ao texto também colaboram com o desenvolvimento de outros subcampos da Organização do Conhecimento, como a Linguística Documental e a Textologia Documental. Pois, são disciplinas ou áreas preocupadas com os estudos das estruturas textuais e seus significados, aplicados para fins de organização e representação da informação.

A Linguística Documental, originalmente proposta por García Gutiérrez, segundo Lara e Tálamo (2007), reúne os estudos sobre problemas de linguagem relacionados com as linguagens documentais no Brasil. Também, contém o ramo da semiótica documental, com estudos das teorias de Peirce aplicados na comunicação de linguagens documentais. Mas, segundo García Gutiérrez (1998, tradução nossa),

> [...] devemos ressaltar a importância da consolidação e expansão da Linguística Documental como disciplina que aprofunda seus fundamentos nos postulados das ciências da linguagem, semânticas e gramáticas, principalmente, diante de intersecções ocorridas com campos afins, como Análise do Discurso, Análise de Conteúdo e, em geral, com as denominadas, e em construção, Ciências Cognitivas.

Por isso, verifica-se que os estudos no âmbito do texto voltados para análise, síntese e representação também constituem essa área, o que permite incluir os estudos sobre o Percurso Gerativo de Sentido na Linguística Documental.

A Textologia Documental, por sua vez, estuda os aspectos discursivos e textuais para organização e representação de conteúdo, relacionados com o ramo da morfologia documental de conteúdo e a forma documental de conteúdo (IZQUIERDO ALONSO, 2004).

Esses estudos de enfoque hjelmsleviano exploram os princípios gerais das estruturas internas das mensagens, a forma e conteúdo dos textos, bem como a relação e função de seus elementos estruturais, entre outros aspectos (IZQUIERDO ALONSO, 2004). Contudo, também são estudos sobre a estrutura e o sentido dos textos, por isso, são paralelos e complementares às demais teorias linguísticas e semióticas afins, que admitem camadas estruturais e de sentido do texto.

Diante do exposto, verificou-se que o Percurso Gerativo de Sentido apresenta caráter semiótico greimasiano e linguístico no âmbito do texto e discurso, também, contribui para o desenvolvimento epistemológico da Análise Documental de Conteúdo e outras áreas afins.

4 Considerações finais

A interdisciplinaridade utilizada pela Análise Documental de Conteúdo de textos narrativos ficcionais, estudada por pesquisadores da UNESP de Marília-SP-Brasil, está centrada na abordagem teórica do Percurso Gerativo de Sentido, proveniente da Análise do Discurso e Semiótica Greimasiana. Também, constitui os aspectos teóricos, metodológicos e epistemológicos que subsidiam a Análise Documental de Conteúdo em procedimentos de análise, síntese e representação do conteúdo desses textos.

O Percurso Gerativo de Sentido permite a identificação do *aboutness* proveniente das regras de significação e estrutura do texto, para fins de construção de índices e indexação de conceitos, enriquecendo conceitualmente e metodologicamente a Análise Documental de Conteúdo e áreas em que mantém interlocução teórica.

Nesse sentido, verificou-se que esses aspectos teóricos interdisciplinares apresentam caráter semiótico greimasiano e linguístico no âmbito do texto e discurso, e colaboram para embasar epistemologicamente também as áreas desenvolvidas com os estudos da Análise Documental de Conteúdo, como o Tratamento Temático da Informação, Linguística Documental, Textologia Documental e o campo da Organização do Conhecimento.

Referências

BEGHTOL, C. Bibliographic classification theory and text linguistics. *Journal of Documentation*, London, v.42, n.2, p.84-113, june. 1986.

FIORIN, J.L. *Elementos de análise do discurso*. São Paulo: Contexto, 2008.

GARCÍA GUTIÉRREZ, A. Elementos de lingüística en sistemas de información y documentación. *Revista Latina de Comunicación Social*, n.7, jul. 1998. Disponível em:<http://www.ull.es/publicaciones/latina/a/66ant.htm>. Acesso em: 07 abr.2013.

GARDIN, J.-C. Document Analysis and Linguistic theory. *Journal of Documentation*, v.29, n.2, june.1973.

_____. *Les analyses de discours*. Switzerland: Delachaux et Niestlé, 1974.

GARDIN, J.-C. et al. *La logique du plausible*. Paris: Maison SH, 1981.

GREIMAS, A.J. *Semántica estructural*. Madrid: Gredos, 1971.

GUIMARÃES, J.A.C. A dimensão teórica do Tratamento Temático da Informação e suas interlocuções com o universo científico da ISKO. *RICI*, v.1, n.1, p.77-99, jan./jun. 2008. Disponível em: <http://seer.bce.unb.br/index.php/RICI/article/view/2761/2331>. Acesso em: 20 fev. 2013.

_____. Abordagens teóricas de tratamento temático da informação. *Ibersid*, Zaragoza, v.3, p.105-117. 2009.

GUIMARÃES, J.A.C. et al. Referentes teóricos em análise documental de conteúdo no ambiente acadêmico espanhol de Biblioteconomia e Documentação. In: ISKO-Español,

10., 2011, Ferrol. *Actas*... Coruña: Universidad Coruña, 2012. p.195-207. Disponível em:<http://www.iskoiberico.org/wp-content/uploads/2014/09/195-207_Guimaraes.pdf>. Acesso em: 27 fev.2015.

GUIMARÃES, J.A.C.; MORAES, J.B.E.de; GUARIDO, M.D.M. Análisis documental de contenido de textos narrativos. *Ibersid*, Zaragoza, v.1, p.93-100. 2007.

HUTCHINS, J.W. On the problem of “aboutness” in document analysis*. Journal of Informatics*. v.1, n.1, p.17-35, 1977.

IZQUIERDO ALONSO, M. Nuevos retos en el análisis documental de contenido. *Scire*, Zaragoza, v.10, n.1, p.31-50, en.-jun. 2004.

LARA, M.L.G.de; TÁLAMO, M.de F.G.M. Uma experiência na interface Linguística Documentária e Terminologia. *DataGramaZero*, v.8, n.5, out.2007. Disponível em:<http://www.dgz.org.br/out07/Art_01.htm>. Acesso em: 12 maio 2012.

MORAES, J.B.E.de. Aboutness in fiction. In: NEELAMEGHAN,A.; RAGHAVAN,K.S.(Org.). *Categories, contexts and relations in KO*. Würzburg: Ergon-Verlag, 2012a. p.242-248.

_____. *A questão do aboutness no texto narrativo de ficção*. Tese (Livre-Docência em Linguística e Documentação)–FFC, UNESP, Marília, 2011.

_____. Perspectivas metodológicas para a identificação do *aboutness* em textos narrativos de ficção. *Scire*, Zaragoza, v.18, n.2, p.67-76, jul./dic. 2012b.

MORAES, J.B.E.de; GUIMARÃES, J.A.C. Análise documental de conteúdo de textos literários. *Scire*, Zaragoza, v.12, n.1 en.-jun. p.71-83. 2006.

VAN DIJK, T.A. *El discurso como estructura y proceso*. Barcelona: Gedisa, 2000.

As linguagens de indexação e a análise de domínio

Vera Lucia Ribeiro Guim
Universidade Estadual Paulista
verinha.guim@gmail.com

Mariângela Spotti Lopes Fujita
Universidade Estadual Paulista
fujita@marilia.unesp.br

1 Introdução

Um dos tratamentos que a ciência da informação aborda é a recuperação e representação da informação. Para tanto é necessário que se utilizem linguagens claras e objetivas designadas por um profissional indexador. Para essas linguagens denominamos linguagens de indexação.

As linguagens de indexação são linguagens construídas artificialmente que visam de maneira sucinta representar o conteúdo do documento. A linguagem enquanto objeto de trabalho traz a tona à necessidade de estudar as ciências que também buscam na linguagem seus objetos de estudo, tais como a linguística, entre outras.

A fim de contextualizar essa questão direcionamos o olhar a analise de domínio, analise esta que abarca as questões históricas, culturais e linguísticas do conhecimento ou da comunidade usuária, acerca da representação da informação nos sistemas ou ambientes informacionais.

Abordada em muitos estudos, a análise de domínio tem sido alvo em diversas pesquisas a fim de fundamentar ainda mais seus temas propostos. Diante dessas considerações procura-se neste estudo discutir uma revisão de literatura e buscar com base nas teorias da análise de domínio os embasamentos teóricos que contribuam e

atendam as linguagens de indexação, sua dinamicidade e representação tendo como fundamento a ciência da informação.

2 Linguagens de Indexação

Imbricada na ciência da informação compreende-se que o entendimento de linguagens de indexação desponta como ferramenta essencial para a produção e organização da informação. Assim, a linguagem se contextualiza dentro do tratamento documental como atividade das unidades de informação cuja intensão é atender as metas e objetivos da organização.

A norma NP 4036 define linguagem de indexação como o conjunto controlado de termos escolhidos numa linguagem natural e utilizados para representar sob forma coordenada o conteúdo dos documentos.

> As linguagens de indexação são instrumentos auxiliares na atividade de representação e recuperação de informações. Atuam nos sistemas de informação orientando o indexador sobre quais os melhores termos para representação do assunto de um documento e orientando os pesquisadores na elaboração das estratégias de busca de informações no sistema. (MOURA, 2002, p.4)

As bibliotecas ou unidades de informação fazem uso das linguagens documentais ou de indexação a fim de descrever o conteúdo dos documentos do acervo com o objetivo de armazená-los e recuperar as informações dispostas nos documentos. Sendo assim as linguagens documentais são divididas em duas classes, as linguagens documentais hierárquicas e as linguagens documentais alfabéticas.

As linguagens documentárias hierárquicas ou pré-coordenadas são notacionais, isto é, os termos se coordenam antes de sua determinação, exemplos de linguagens documentais hierárquicas são os sistemas classificatórios, como CDU (Classificação Decimal Universal) e CDD (Classificação Decimal de Dewey).

Nas linguagens documentárias alfabéticas ou pós-coordenadas, os termos se coordenam após sua determinação. E são formadas pelos vocabulários controlados como os tesauros ou cabeçalhos de assunto.

A fim de uma utilização ideal das linguagens documentais torna-se necessário interligar a linguagem natural utilizada pelo usuário com a linguagem designada pelo sistema de recuperação utilizado, uma vez que em determinas regiões do país uma palavra buscada pode ter outro significado. Assim a fim de facilitar o processo de comunicação entre o usuário e o sistema são estruturados instrumentos de controle para os termos a serem utilizados, portanto é necessário que exista um pleno domínio na linguagem utilizada de maneira que os termos possam ser conhecidos e facilmente utilizados pelo indexador no sistema e consequentemente pelo usuário.

De acordo com Vale (1987), a escolha de uma linguagem de indexação é um fator de grande importância para eficácia de um sistema de recuperação da informação, logo tal escolha necessita considerar os objetivos do sistema, o tipo de usuário e a abrangência ou a especificidade do assunto a ser tratado.

2.1 Linguagens Pré-Coordenadas e Pós-Coordenadas

Ao organizarem as informações seguindo uma estrutura hierárquica e sistemática, as classificações consideradas linguagens pré-coordenadas, pois os conceitos são representados pelo indexador no momento da representação, “são instrumentos por excelência usados na organização do conhecimento, primeiro numa arrumação física, depois, e em paralelo, na organização lógica e sistemática do conhecimento em catálogos” (SIMÕES, 2008, p.36).

A estrutura oferecida pelas classificações permite que as informações sejam agrupadas de forma coerente, contextualizada e lógica.

As Linguagens pós-coordenadas segundo Lancaster (2002, p. 22) buscam,

1- Facilitar a representação consistente das matérias por parte dos indexadores e utilizadores que recuperam, evitando a dispersão dos elementos relacionados.

2- Facilitar a realização de uma pesquisa ampla sobre um assunto enlaçando os termos com relações paradigmáticas ou sintagmáticas.

As listas de cabeçalho de assunto e os tesauros continuam a ser as linguagens controladas pós-coordenadas mais utilizadas na representação e recuperação da informação.

2.2.1 CDD e CDU

Os sistemas de classificações que mais se popularizaram, datam do século XIX, a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU). Esses esquemas de classificação bibliográfica buscam mapear o conhecimento em suas diferentes áreas dispondo os assuntos com base nas divisões hierárquicas partindo do geral para o mais específico.

Para Cintra, (2002, p.44),

> Nos sistemas de classificação bibliográfica, a estrutura hierárquica é dada pela notação (decimal, no caso da CDD e da CDU). O vértice das cadeias hierárquicas é constituído por disciplinas convencionais que se subdividem sucessivamente. A indicação dos assuntos é feita por meio da notação numérica ou alfanumérica, conforme o tipo de sistema.

Atualmente em sua vigésima segunda edição, a CDD é considerada o sistema de classificação mais utilizado no mundo adotada em mais de 135 países e traduzido para mais de trinta idiomas.

A Classificação Decimal Universal (CDU) foi criada em 1905, construída pelos belgas Paul Otlet e Henri La Fontaine. Os autores adaptaram o sistema da CDU creditando a ela um caráter mais enumerativo e linear, o que ocasionou um avanço na representação de assuntos sejam eles mais simples ou mais complexos por meio dos mecanismos de combinação.

2.2.2 Tesauro

Para a UNESCO (1973, p.6) o tesauro,

> É um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente cobrindo um domínio específico do conhecimento [...] É um dispositivo de controle terminológico usado na tradução da linguagem natural dos documentos, dos indexadores ou dos usuários numa linguagem do sistema (linguagem de documentação, linguagem de informação) mais restrita.

Segundo seus principais utilizadores os tesauros adequam-se a áreas especializadas, em geral são mais fáceis de utilizar, exigem menos tempo ao indexador e eram mais rápidos na recuperação da informação. Portanto seguem os princípios da pós-coordenação.

Segundo o IBICT (1984, p. 1-2), as principais finalidades de um tesauro são:

a) controlar os termos usados na indexação mediante um instrumento que traduza a linguagem natural dos autores, usuários e indexadores, para uma linguagem mais controlada;

b) uniformizar, mediante esta linguagem documentária, os procedimentos de indexação de profissionais em uma instituição ou numa rede cooperativa;

c) limitar o número de termos necessários à explicitação dos conceitos expostos pelos autores de uma área;

d) auxiliar a tarefa de recuperação da informação, fornecendo termos adequados para a estratégia de busca.

Os tesauros, exemplos de vocabulário controlado, têm como principal função ter o controle terminológico do vocabulário utilizado em uma área de conhecimento especifica.

2.2.3 Cabeçalhos de assunto

Os cabeçalhos de assuntos buscam representar os assuntos em cabeçalhos estruturados. No entanto é considerada umas das linguagens mais custosas, pois exigem do indexador muito mais tempo, porém a especificidade e precisão da informação são mais garantidas possibilitando também uma maior precisão na descrição dos assuntos complexos e permitem uma percepção mais imediata, por parte do usuário.

> Os cabeçalhos de assunto representam uma linguagem muito estruturada e pré-coordenada, oferecendo ao pesquisador possibilidades limitada de

> modular a sua pesquisa. Entretanto, existem hoje vocabulários que se intitulam listas de cabeçalho de assunto e são utilizadas em sistemas pós-coordenadas, como é o caso do MESH (Medical Subject Headings). (CESARINO; PINTO, 1978)

Sendo assim como toda linguagem as listas de cabeçalho de assunto também necessitam de atualizações uma vez que quanto maior o volume de documentação maior será o grau de especificidade e para isso é necessário um controle adequado e de qualidade buscando beneficiar tanto o indexador, a instituição e o usuário desse tipo de linguagem.

As linguagens documentais se fazem necessárias, pois buscam conciliar o vocabulário a ser utilizado pelo sistema, pelo indexador e pelos usuários a fim de que diferentes termos sejam usados para assuntos idênticos evitando assim que documentos similares acabem separados.

Dessa forma as linguagens documentais acabam por contribuir a uma delimitação do domínio dos campos de estudo. Cintra et.al. (2002) afirmam que “isto é possível porque para cada unidade preferencial integrada numa linguagem documental é necessário que corresponda um conceito ou noção [...] uma vez que só a organização nocional de uma área permite a utilização de instrumentos eficazes para o tratamento e recuperação da informação”.

A fim de avaliar as linguagens utilizadas pelos sistemas de recuperação busca-se analisar qual o domínio que a linguagem pertence, a fim de verificar se a linguagem é pré ou pós-coordenada assim como sua representação e elaboração.

3 Análise de Dominio

A análise de domínio procura identificar e analisar áreas específicas a fim de construir um sistema de informação, buscando também construir modelos de representação da informação de acordo com os campos específicos do conhecimento e suas características.

Para Danuello (2007) a análise de domínio “é um processo pelo qual se pode verificar o que é importante ou significativo sobre algum campo do esforço humano, identificando elementos que permitam analisar um contexto cientifico”. Lorenzon (2011)

afirma ainda que ela “é feita com base nas condições históricas, culturais e linguísticas das comunidades discursivas (usuários) que direcionam a construção de sistemas de classificação e tesauros para a representação da informação em seus ambientes informacionais”.

Por não ser possível prever as necessidades informacionais dos usuários e ter consciência que o indexador pode se deparar com vários domínios de conhecimento, segundo Hjørland (2002, p.422) a análise de domínio se volta para o estudo e a estruturação de áreas de assuntos, da mesma forma, para os sistemas de classificação, porém, com a contextualização social dos conceitos, o que reflete no critério de que um documento jamais deve ser analisado isoladamente.

Logo, a fim de analisar um domínio procura-se identificar, explicar e prever todos os fatos que possam ocorrer em determinado conjunto de termo dentro desse domínio. Isto para que posam servir como uma fonte de referencia, quando ambiguidades surgirem, como um repositório de conhecimento comum, auxiliando de forma direta na comunicação, no aprendizado e no reuso desses termos e conceitos em qualquer nível de abstração.

Hjørland (2002) definiu que para a organização do conhecimento é necessário que a representação seja feita considerando conhecimentos sociais, culturais, linguístico e cognitivo de determinadas áreas especificas do conhecimento.

Autores como Hjørland e Tennis entre outros discutem a analise de domínio e as teorias dos autores em muito contribuíram para a consolidação deste modelo de analise na ciência da informação.

Dessa forma, acredita-se que com a contribuição teórica metodológica dessa abordagem torna possível contextualizar a pesquisa no âmbito de domínio específicos das linguagens de indexação.

4 Linguagens de Indexação e Analise de Dominio

Dessa forma, Hjørland (2002, p.430) conclui que a indexação e recuperação da informação são sempre específicas, e recomenda que pesquisadores e profissionais da área de indexação e recuperação atendam as demandas de diferentes domínios em

sistemas de organização e recuperação de documentos ressaltando que um enfoque mais forte em diferentes domínios poderá tornar nossa área mais realista e mais relevante em diferentes ambientes ou contextos.

Independente da linguagem escolhida pela instituição, no momento em que o indexador decide qual linguagem utilizar há um ponto de partida para sua criação e pesquisa básica e para tanto é necessário que se volte o olhar a concepção da analise de domínio que busca seguir os métodos das ciências sociais, segundo seus atores e o coletivo externo as posturas subjetivas.

> Uma forma diferente de análise de domínio surge a partir da necessidade de compreender um domínio para se construir ou revisar um sistema de informação. Por exemplo, os editores da Classificação Decimal de Dewey a revisam com o objetivo de atender às mudanças no domínio representado por esta classificação universal. (TENNIS, 2012)

Assim como para uma pesquisa, a construção de um sistema de informação em muito se baseia na analise de domínio, pois esta,

> [...] é uma metodologia para construção de modelos de representação da informação a partir da investigação de determinadas características de domínios específicos do conhecimento: a identificação de condições culturais, históricas e linguísticas que imponham exigências particulares para a construção de modelos de domínio tais como sistemas de classificação ou tesauros. Compreende também um ponto de vista epistemológico para identificar os paradigmas científicos e técnicos, abordagens de pesquisa e interesses de conhecimento nos domínios cobertos. (NOVELLINO, 1996)

Sendo assim percebe-se que as diferentes linguagens e suas metodologias e métodos traduzem-se enquanto domínios uma vez que permite a construção de sistemas e de pesquisas tendo em vista seus objetivos e fundamentos.

5 Considerações

Diante de diversas aplicações e adaptações os estudos da analise de domínio estão cada vez mais relacionados e interconectados com as mais distintas áreas do

conhecimento. A sua relevância frente às linguagens está firmada e inferida em seus estudos.

As funções e ações da analise de domínio para a indexação é de clara importância e condicionalidade. Cada aplicação aponta seu fundamento e abordagem, assim como sua extensão e exclusão visando contribuir para o aprofundamento das questões de representação e tradução e valor da analise de domínio, ponderando seu alcance e escopo em relação às linguagens de indexação.

Sendo essas, tidas como fundamentais para uma descrição e indexação adequada aos sistemas e meios informacionais, pressupõe-se que as experiências a que as linguagens de indexação são acometidas possam sugerir a cultura, o meio ambiente, a educação e até mesmo o emocional que a analise de domínio utiliza para os fins de pesquisa.

Infere-se também que a forma estruturada das linguagens, suas traduções e representações nas unidades de informação influenciam diretamente nos propósitos da organização.

Considerando isso, os autores, citados neste estudo, reforçam a necessidade de uma linguagem de indexação estruturada a partir da profundidade teórica fornecida pela analise de domínio e que busca ser entendida como ferramenta imprescindível de modo a promover e agregar valores, eficácia e consistência a informação.

Referencias

CESARINO, M.A. N; PINTO, M.C.M.F. Cabeçalho de assunto como linguagem de indexação. **R. Esc. Bibliotecon.** Belo Horizonte: UFMG, v.7 (2) p.268-88, Set. 1978.

CINTRA, A.M. M (et.al). **Para entender as linguagens documentárias**. 2. ed. rev.e ampl. - São Paulo: Polis, 2002

DANUELLO, J. C. **Produção científica docente em tratamento temático da informação no Brasil**: uma abordagem métrica como subsídio para a análise de domínio. Marília: UNESP, 2007. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação - Universidade Estadual Paulista (UNESP).

HJØRLAND, B. Domain analysis in information science: Eleven approaches – traditional as well as innovative. **Journal of Documentation**, v.58, p.422-462, 2002. Disponível em: www.db.dk/bn/publikationer/Filer/JDOC_2002_Eleven_approaches.pdf

IBICT. Diretrizes para elaboração de tesauros monolíngues. Brasília, 1984.

LANCASTER, F.W. **El Control del Vocabulario en la Recuperación de Información**. 2. ed. Valencia, 2002.

LORENZON. E.J. **Análise de Domínio para Avaliação de Tesauros**: Uma experiência com a cadeia produtiva do calçado no Brasil. Marília: UNESP, 2011. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação - Universidade Estadual Paulista (UNESP).

MOURA, M.A.; SILVA, A.P; AMORIM, V.R. A concepção e o uso das linguagens de indexação face às contribuições da semiótica e da semiologia. **Informação e Sociedade: Estudos.** V.12, n.1, 2002

NOVELLINO, M.S.F. Instrumentos e metodologias de representação da informação. **Informação e Informação**, Londrina, v.1, n. 2, p. 37-45, jul./dez. 1996.

SIMÕES, Maria da Graça. **Da abstração à complexidade formal**: relações conceptuais num tesauro. Coimbra: Almedina, 2008

TENNIS, J.T. **Two axes of domains for domain analysis.** Knowledge Organization, 30(3/4), 2003, p.191

TENNIS, J.T. Com o que uma análise de domínio se parece no tocante a sua forma, função e gênero? **BJIS**, Marília (SP), v.6, n.1, p.3-15, jan./jun. 2012.

VALE Eurides A. do. Linguagens de indexação. In: SMIT, Johanna W. (Coord.). **Análise documentária**: a análise da síntese. Brasília: NCT – CNPQ, 1987. Cap. 1, p.12 – 26.

Reader-Interest classification as a method of stock control: The McClellan Legacy

Daniel Martínez-Ávila
Universidade Estadual Paulista
dmartinezavila@marilia.unesp.br

M P Satija
Guru Nanak Dev University - India
satija_mp@yahoo.com

1 Introduction

A "reader-interest classification" is "a classification designed to serve the immediate needs of the targeted users. Such systems may violate the filiatory sequence to bring together disparaged subjects needed by a user group. Indeed these are useful in mission oriented or multidisciplinary subjects. In a commerce college, e.g., it may be more pragmatic to place commercial law with commerce at 380. It is true to say that reader's interest classification adopted so far are not satisfactory in the long run and sometimes correspond to ephemeral vogues. It reflects a middle level of ambition in knowledge organisation. It is a compromise between ad hoc classification and rigorously scientific classification" (Satija 2004, p.182). Although not exempt from controversy, the study of reader-interest classifications has been recently reviewed (e.g., Satija 2009) and revived in the light of the recent cases of adoption of the book industry classification BISAC (e.g., Martínez-Ávila and San Segundo 2013; Martínez-Ávila et al. 2014). In this paper, we review the classic literature on reader-interest classifications to reveal one of the key and most remarkable claims behind these cases: the use of reader-interest classification as a method of stock control.

2 Reader-interest classifications and stock control

When dealing with physical collections (as did all cases which this work has examined in the literature) one of the main advantages/characteristics of reader-interest classifications was their relationship with stock control, stock management, and arrangement. Stock control, generally speaking, is the process of maintaining an appropriate level of stock to meet the demand while keeping the costs to a minimum. Stock control was claimed to be one the main goals of many reader-interest classification projects such as in libraries of Cambridgeshire (Chandler 1982), Cheshire (Astin 1982), East Sussex (Donbroski 1982) and Surrey (Betts 1982). All these libraries claimed to implement the reader-interest classifications for a variety of purposes indirectly related to stock control. In Cambridgeshire the reader-interest classification was claimed to be adopted to improve self service, in East Sussex for the development of the collection, and in Surrey as a new policy of provision of stock to identify the demand. In all these cases stock control was considered a vital part of the projects and therefore linked to a variety of other aspects and advantages. According to Ainley and Totterdell, "the relationship of stock control to categorisation is vital - in fact the two are really inseparable" (Ainley & Totterdell 1982, p.126). It might be added that they are inseparable because they are mutually interdependent: reader classifications affected stock control, and stock control was usually one of the main advantages/desired goals of reader-interest classifications.

In any case, the relationship between these two elements seems based on the assumption that reader-interest classifications not only meet user's needs, but enable the most relevant areas of the collection to be identified. Once these areas of stock are physically isolated, it is then possible to identify reader's interests according to the demand/supply relation. This use of reader-interest classification allows better policies of book selection, arrangement, and weeding, while keeping the most adequate levels of stock for the user. In addition, the relationship of demand/supply and the readers' interest was also found to be linked to the concept of stock management.

Stock management, as defined by Astin, is "the committed operation of a system of procedures for the informed allocation and redistribution of stock controlled by data as to potential and actual use against predetermined targets" (Astin 1982, p.15). Even when not serving a primary purpose, stock management was claimed to be an advantage of reader-interest classifications -as was discovered with the first reader-interest

classification put into practice at Detroit Public Library. Observing the effects of this new system upon the collection, Rutzen declared that "[the reader-interest classification] affects book selection. We are beginning to be aware that some books generally purchased by branches operating under the old system may not be included by those using this classification, and vice versa" (Rutzen 1952, p.482).

Other libraries adopting a reader-interest classification also noticed the practical benefits of physical display or visualization. Andrew Miller, at Glasgow City Libraries, claimed that "arranging the stock by categories does increase demand and highlights deficiencies in stock holdings" (Miller 1992, p.132). Sapiie pointed out that: "it was found that the reader-interest classification was also useful for the meaningful collection of data and assessing and editing stock because it provides a visual means to determine the size of various collections, to identify low use or surplus stock and areas for future stock provision. The reorganization process provides an excellent opportunity for stock weeding. Since the stock is physically divided into manageable groups,a reader-interest classification provides an easy way to determine how many books of a particular genre a library owns. Future stock selection is facilitated because reader-interest classification gives a clear indication of reading preferences" (Sapiie 1995, p.151).

The concept of stock management was closely related to the control of stock levels, shelf stock and the proper number of books needed for reader-interest classifications. For Douglas Betts, librarian at Surrey, reader-interest classifications were linked to the adequate control of stock levels: "some form of categorisation is an important pre-requisite for constructive thinking about stock levels" (Betts 1982, p.70). Yet at the same time it was again difficult to link stock control solely to the classification system; at Surrey, stock reduction was a basic requirement for many measures related to guiding and space, such as "face-on" and integration of all sizes books. Importantly, stock reduction was also related to the need to maximize resources and loans. As pointed out by Alex Wilson (and later by Ainley and Totterdell) a shelf arrangement based on reader interest categories should be "a first call on limited resources" (Wilson 1978). The point was to maintain stock in certain areas at the correct levels -arguably at a minimum- in order to serve a significant number of users and to thus to save costs and to avoid distraction to the reader. After determining which areas of interest (or categories in the reader-interest classification

schemes) these should be, the collection was to then focus on developing these areas in order to improve the indicators of library performance and to save readers from information overload.

3 The McClellan formula and reader-interest classifications

The principles of these methods of stock control, in relation to reader-interest classifications, were first developed by Archibald William McClellan, Chief Librarian of Tottenham, who devised a system to determine some kind of proportional representation of stock based on a regular (3-4 times a year) count of books on shelves and on loan. McClellan is also credited as the father of two other concepts closely related to reader-interest classifications: the integration of reference and lending stock and the creation of separate 'popular' sections (Ainley & Totterdell 1982, p.7). According to Ainley and Totterdell (p.127), "the idea was to break away from equal funding for each category as some attract only a few readers but others attract hundreds." McClellan's formula of desired stock was used to determine the right number of books that should be provided/maintained for each category of the reader-interest classification. This formula involves variables such as the total shelf capacity of the library and the highest number of items on loan per category over a period of time (commonly the previous year). This formula was mathematically represented as follows (McClellan 1978, p.8):

$$Ts = S\left(\frac{\sqrt{L1}}{R}\right)$$

where;

Ts = Target shelf stock for an Interest Category

S = Total circulation shelf capacity of library

L1 = Highest number of books on loan (the Loan Component) as shown by the issue "counts" on the previous year

R = Total of all square roots for all Interest Categories 1 to n (i.e. L1 to Ln)

On the other hand, McClellan's approach was also object of criticism. Ainley and Totterdell listed some of the disadvantages as follows: It was essentially concerned with the equitable allocation of existing resources, rather than with the ideal level of provision;

unsatisfied demand was not measured -this could lead to a vicious circle, where unsatisfied demand leads to a reduction in borrowing, which leads to a reduction in titles made available on shelves, resulting in more unsatisfied demand; other immeasurable factors -quality, condition, lighting, etc.- might also lead into the same downward spiral; and the system was devised to be operated with the Browne issue method which made tremendous demands on staff time both during the counts and for the necessary calculations which follow. Currently, this method is no more used in automated charging systems.

However, despite this criticism, McClellan's formula was used in several libraries as a method for stock control. Cheshire libraries, that adopeted reader-interest classifications, were among the library systems influenced by McClellan. Here the most important motive for change was the need to assess performance of stock against demand -that is, keeping a minimum shelf stock in every identified topic of interest for a significant number of users. In 1975 most proposals for a categorized bookstock in Upton (Cheshire) were based on such factors: the arrangement of books according to majority user interests (and a better understanding of the stock/issue relationship); easier identification of dead and surplus stocks allowing transfers and control on future provision; the facility to monitor 'saturation' and 'starvation' effects within categories; easier shelving of returned books and best housekeeping; and a better understanding of the real relationship between supply and demand and a more flexible response to user needs (Astin 1982, p.14). Indeed, according to Astin, the reader interest classification and the McClellan formula in Cheshire were weighted to protect minority interests (Astin 1982, p.17). However, such a statement seems to contradict the vision that the library had of itself and the project: "Libraries with stock arranged in an order which matches the selection methods of the majority of our borrowers. Libraries guided to a uniform and high standard of presentation" (Astin 1982, p.18). In general terms, it would seem difficult to protect the interests of the minority while matching the selection methods of any majority.

Stock management thus seems to have been inextricably linked to meeting user's needs. When East Sussex County Library, for example, adopted a categorization system it declared (i) that the arrangement of stock in broad interest categories was more suited to the needs of most users than traditional classified arrangements; and (ii) that this

categorization of stock should enable staff to maintain a more effective stock provision policy (Donbroski 1982, p.79). Upon concluding the reader-interest classification project at Peacehaven and Hangleton libraries, librarians found that "deficiencies in the balance of provision could be identified through the categorised arrangement. […] even without a fully systematic method of stock provision, reader interest arrangement of stock can, at the expense of very little effort, provide a clear guide to the relationship between levels of provision and use" (Donbroski 1982, p.83).

Improvements in stock control also appeared to be linked to the lack of order within categories -- an aspect commonly associated with reader-interest classifications. "Indeed librarians at both libraries have suggested that the less precise order has positive benefits in promoting greater stock knowledge amongst staff at all levels" (Donbroski 1982, p.85). In this vein, management of stock on the shelves and how this relates to user's interests was also informed by the debate over accessibility and the role of purposive and non-purposive readers found in reader-interest classifications discourses.

Also relevant -- and detailed in the literature- - was the relationship between the proper number of physical stock and a good performance in retrieval. The proper amount of stock for a good retrieval is related to the coverage (in the reader-interest classifications) of the interesting areas of stock. First, it is assumed that lower the number of stock in a library, the higher the performance in finding and retrieval by browsing. Assuming that not all the collection would be equally useful for all users, there is also the need to maintain a collection which is of interest to a particular community or user. As such, it is necessary to decide which community or users should benefit in order to keep a balanced number of stock and to improve 'retrievability.' For Astin, this is similar to the concept of “substitute borrowing” (Astin 1982, 15): "surveys have revealed that the greater majority (85%+) of public library borrowers do not require specific authors or titles but want 'good reads' in topics of related interest: i.e. they are satisfied by an adequate range of general choice in which any one title is an acceptable substitute for any other." These results were supported by a user survey carried out at Hillingdon library in the UK (Donbroski 1982, p.81), which claimed to show a high amount of substitute borrowing, and by Southwick Branch Library which found that "fewer books did not mean an immediate drop in issues" (McCarthy 1982, p.88).

Finally, Surrey Library also proposed that small and medium libraries should have very little "permanent" stock (Ainley & Totterdell 1982, p.128), since this would make more titles available (by circulating other libraries' stock) and would mean greater exposure for the "low demand" categories. The argument here seems to be that by covering categories or areas of interest for users with any kind of title, libraries are able to address the user's interests independently of the titles and the quality of those books. Depending on the philosophy of each reader-interest classification project, there might be cases where this is considered an advantage for libraries.

4 Conclusions

It can be inferred from the literature that reader-interest classifications, whose focus is claimed to be on the interest of the readers, can be a method to determine the areas of the collection (i.e. the areas of the universe of knowledge to be organized with the classification system) that the users are more interested in, and therefore that must be provided with more books. While reader-interest classifications are presented as systems that collocate r distributed areas of common interest, the suggestion of using them for stock control also introduces one more variable: the degree of interest (intension) in addition to the association of subjects of common interest (extension). For example, different topics might be gathered together in a class as they might respond to a common interest of a person, but different classes (of common interest) might be of the interest of that person at different degrees. However, it is not clear what aspects of the extension of the interests, the intension of the interests, and the interests of the readers are affecting what and in which order, especially in a collection that is going to privilege the intension of the interest to determine the extension (the classes). The risk of influencing on the interests of the readers with the popularity of several topics is a shortcoming that has not been overcome by the stock control function. In this vein, it seems that aspects such as display, signaling, lighting, etc. that affect the outcome of reader-interest classifications might be behind the true factors affecting the stock control function too.

References

Ainley, P.; Totterdell, B.. 1982. *Alternative Arrangement: New Approaches to Public Library Stock*. London: Association of Assistant Librarians.

Astin, J. 1982. Cheshire: Alternative Arrangement and Beyond. In Patricia Ainley, and Barry Totterdell, eds., *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, pp. 13-33.

Betts, D. 1982. Reader Interest Categories in Surrey. In Ainley, P.; Totterdell, B. eds., *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, pp. 60-77.

Bragg, S. 2015. *Inventory Management: Second Edition*. Centennial, Colorado: Accounting Tools

Chandler,D. 1982. Self-service Libraries:Providing for the Smaller Community in Cambridgeshire. In: Ainley,P.; Totterdell, B. ed. *Alternative Arrangement:New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, p. 47-59.

Donbroski, L. 1982. Categorisation at East Sussex County Library. In Patricia Ainley, and Barry Totterdell, eds., *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, pp. 78-85.

Martínez-Ávila, D, San Segundo, R.. 2013. Reader-Interest Classification: Concept and Terminology Historical Overview. *Knowledge organization* 40: 102-14.

Martínez-Ávila, D, San Segundo, R.; Olson, H.A.. 2014. The Use of BISAC in Libraries as New Cases of Reader-Interest Classifications. *Cataloging & classification quarterly* 52 no.2: 137-55.

McCarthy, A. 1982. Burning Issues: Stock Appeal in Sunderland. In Ainley, P., Totterdell, B. eds., *Alternative Arrangement: New Approaches to Public Libraries Stock*. London: Association of Assistant Librarians, pp. 88-100.

McClellan, A W. 1978. *The Logistics of a Public Library Bookstock*. [Harrogate]: Association of Assistant Librarians.

Miller, A. 1992. Alternative Arrangement in Glasgow City Libraries. Public *Library journal* 7 no.5: 131-3.

Rutzen, R. 1952. Shelving for Readers. *Library journal* 15: 478-82.

Sapiie, J. 1995. Reader-interest Classification: the User-friendly Schemes. *Cataloging & classification quarterly* 19 no.3/4: 143-55.

Satija, M P. 2004. *A Dictionary of Knowledge Organization*. Amritsar: Guru Nanak Dev University.

Satija, M P. 2009. Users Interest Classification. *SRELS journal of information management* 46 no.4: 353-4

Wilson, Alex. 1978. The Threshold of Choice. In T*he Public Library and Its Users. Proceedings of the Public Libraries Group Weekend School, University of Warwick 25 - 28 March*. London: Library Association, Public Libraries Group.

Terminologia, exaustividade e especificidade: uma relação conceitual

Isadora Victorino Evangelista
Universidade Estadual Paulista
isadora.biblio@marilia.unesp.br

Walter Moreira
Universidade Estadual Paulista
walter.moreira@marilia.unesp.br

João Batista Ernesto de Moraes
Universidade Estadual Paulista
prof.joao@gmail.com

1 Introdução

A indexação e a construção de terminologias são tarefas que podem ser agrupadas sob o denominador comum da identificação, descrição e controle das relações conceituais que os termos estabelecem entre si. Desse modo, ambas também se preocupam com a análise conceitual enquanto recurso para a efetiva compreensão dos conceitos em si mesmos.

A indexação, um dos processos da representação temática, é realizada com o propósito de atender às necessidades de informação de um determinado grupo de usuários. Isso implica dizer é preciso relacionar a análise do documento em si, isso é a identificação e a representação de seu conteúdo, com os interesses do grupo a que se destina. A atribuição de termos que sejam representativos dos conteúdos dos

documentos e que sejam, ao mesmo tempo, igualmente representativos nas linguagens de especialidade dos grupos de usuário, é uma decisão terminológica.

A indexação possui duas características que lhe são orientadoras, a exaustividade e a especificidade. De modo sucinto, a primeira, diz respeito à extensão com que um assunto será coberto enquanto a segunda, diz respeito à precisão com que os termos irão representar os assuntos tratados no documento. Por sua vez, a terminologia se caracteriza por ser o estudo dos termos que são representativos de determinada especialidade, de um domínio do conhecimento. É nos estudos terminológicos que se concentram os estudos relativos aos termos, a abrangência com que um domínio deve ser pesquisado para que sejam criadas as representações das suas linguagens de especialidade.

O presente artigo tem por objetivo verificar as relações entre a terminologia teórica e os conceitos de exaustividade e de especificidade conforme suas expressões na literatura periódica brasileira. O traballho se justifica pela contribuição à literatura das duas áreas e contribui para que esses dois universos possam dialogar ainda mais.

2 Indexação

A organização e representação do conhecimento é uma das áreas da Ciência da Informação que tem por interesse os subsídios teóricos que se relacionam ao tratamento da informação (BARITÉ, 2001) e ainda, com os procedimentos e sistemas de representação da informação, de maneira a utilizá-los de forma eficaz (HJØRLAND, 2008). Para tanto, alguns processos são necessários, como a indexação.

A indexação diz respeito à elaboração de índices, procurando descrever tematicamente o documento, por meio de termos definidos a partir de um vocabulário controlado, valendo-se de representações próximas do conteúdo documental. Collison (1972) nos traz duas assertivas relacionadas à indexação: seu principal propósito é colocar ao alcance de qualquer leitor, de forma rápida, as informações contidas no recurso informacional e ainda, este não é um processo mecânico, é necessário que em todas suas etapas haja reflexão por parte do indexador.

Autoras como Fujita e Rubi (2003) e Boccato (2009) definem a indexação como um dos principais processos na biblioteca, por ser uma ferramenta importante na aproximação entre o usuário e o documento representado. Para Fujita e Rubi (2003), a indexação “condiciona” os resultados no momento da recuperação da informação podendo, portanto, trazer reflexos positivos ou negativos, isso é, pode tanto auxiliar na localização de um determinado item como pode efetivamente escondê-lo no sistema.

De acordo com a norma ANSI/NISO Z39.19, a indexação é o processo de atribuição de termos para a descrição do conteúdo dos documentos. Ainda de acordo com essa norma, a indexação abrange todos os procedimentos de seleção e organização de termos e requer procedimentos humanos, embora possa haver também, processos computacionais envolvidos.

Já na norma ABNT/NBR 12676/1992, a indexação é definida como o “ato de identificar e descrever o conteúdo do documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação” (ABNT/NBR 12676/1992, 1992, p. 2).

Lancaster (2004) diz que não existe uma maneira correta de indexação, mas que uma boa indexação irá considerar além dos assuntos contidos no documento, o porquê daquele recurso ser importante para a comunidade usuária. Comunidades usuárias que possuem interesses informacionais diferentes, buscam de forma diferente, por termos diferentes e, portanto, o indexador deve estar atento que não existe um conjunto único e correto de termos representativos.

A indexação possui duas medidas para auxiliar sua eficácia: a exaustividade a especificidade, as quais serão abordadas na sequência.

3 Exaustividade e Especificidade

O processo de seleção de conceitos é intrínseco à indexação, nesse caso é necessário determinar em um universo de conceitos quais serão os mais representativos daquele documento e deve-se levar em conta a finalidade da documentação e dos usuários. As características que melhor convêm nesse sentido, são a exaustividade e a especificidade (ABNT/NBR 12676/1992, 1992).

A exaustividade é definida como os limites de abrangência e o número de conceitos representativos pelos termos conferidos a um documento. Esses limites de abrangência não devem ser vistos de maneira restritiva, uma vez que um documento utilizado por um grupo de usuários pode também, ser utilizado por outro grupo com interesses diferentes (ABNT/NBR 12676/1992, 1992; LANCASTER, 2004).

Para Maturana (1992), a exaustividade é a habilidade de extensão com que os diferentes temas de um documento serão abordados, identificados a partir da análise documentária pelo indexador e, posteriormente, traduzidos a partir de uma linguagem documentária quando passa a ser, então, termos representativos.

Quanto à especificidade, podemos dizer que diz respeito ao nível de precisão que os termos irão ter; o quão específico serão os assuntos cobertos. Para Lancaster,

> Enquanto a exaustividade diz respeito ao âmbito de cobertura, a especificidade refere-se à profundidade de tratamento do conteúdo. Estabelece-se a exaustividade como uma decisão da política de indexação, enquanto a especificidade é uma propriedade do vocabulário adotado na indexação (LANCASTER, 2004, p. 203).

De acordo com a norma NISO TR02:1997 (1997), a especificidade é a proximidade entre o termo definido e o assunto do documento. Essa norma apresenta um exemplo: “caminhões estilo pick-up” pode ser usado para aquele tipo específico de caminhão enquanto que, o termo “caminhão” seria representativo a todos os tipos de caminhão. Ainda de acordo com a norma, a exaustividade e a especificidade podem ser combinadas para determinar a profundidade da representação.

Podemos afirmar que esses dois processos são complementares na indexação, para que o conteúdo temático do documento seja abrangido completamente e que os termos que irão representa-lo, sejam precisos. Na próxima seção, são apresentados alguns pressupostos teóricos sobre os estudos terminológicos para posteriormente, analisarmos como os conceitos de exaustividade e especificidade são relacionados a esse universo.

4 Estudos Terminológicos

Os estudos terminológicos se iniciam com Eugen Wuster, na década de 30, considerado o fundador da terminologia moderna e um dos principais representantes da escola de Viena. Wuster define a terminologia diante de três assertivas: um arranjo de conceitos próprios a um domínio especializado e de suas denominações, a lexicologia especializada de um domínio e a *teoria geral da terminologia*, quando afirma que em terminologia são princípios comuns que são os mesmos em domínios variados, independente da língua (FAULSTICH, 2001).

Outra pesquisadora relevante nessa área é Maria Tereza Cabré, que define terminologia como sendo “a disciplina cujo objeto é o estudo e a compilação dos termos especializados” (CABRÉ, 1993, p. 21). Para a autora, embora o estudo da terminologia, como conhecemos hoje, seja relativamente recente, iniciando-se em meados da década de 30, basta nos recordarmos dos trabalhos de Lavoisier em química e Lineu em botânica, para vermos que há o interesse nas denominações dos conceitos científicos por parte dos especialistas.

Para a autora, a preocupação dos linguistas nos últimos anos concentra-se em construir uma teoria que lide com os princípios das línguas reais e possíveis e não se atenta aos aspectos comunicacionais dos instrumentos. Em uma concepção comunicativa da linguagem, a terminologia poderia auxiliar o trabalho dos linguistas.

Nesse sentido, Cabré desenvolve, juntamente com o grupo de pesquisadores da Universidade Pompeu Fabra, na Espanha, a Teoria Comunicativa da Terminologia, na qual se apresentam fundamentos que são estruturados a partir do enfoque comunicativo das linguagens especializadas, como uma preocupação maior do que os propósitos normalizadores (KRIEGER, 2000).

Na década de 90, os estudos acerca das terminologias são intensificados, considerando a complexidade que envolve seu funcionamento, recebendo contribuições de áreas como a sociolinguística, tradutologia e ainda, do campo da inteligência artificial (KRIEGER,2000).

Em um estudo realizado por Cervantes, Fujita e Rubi (2008), as autoras consideram de grande importância o estudo dos conjuntos terminológicos nas áreas da ciência, sendo importante essa preocupação com as linguagens especializadas, já que

além de determinar o domínio, elas têm a capacidade de auxiliar na transmissão de conhecimento.

Krieger (2000) aponta que as terminologias, enquanto sistemas linguísticos, não possuem neutralidade, já que são advindas de linguagens naturais e expressam conhecimentos especializados e ainda, não são absolutamente artificiais, uma vez que buscam se construir a partir da linguagem, para posteriormente identificar conceitos precisos em diferentes campos do conhecimento, a partir de termos, evitando a polissemia e ambiguidades.

5 Procedimentos metodológicos

A pesquisa proposta possui caráter exploratório, uma vez que se apresenta enquanto um estudo preliminar para que a ideia do objetivo seja apresentada e e para que se formulem hipóteses posteriormente. Ela também se classifica como qualitativa, considerando que o objeto de pesquisa foram artigos que comprovassem a relação da terminologia com os conceitos de exaustividade e especificidade.

Para estabelecer o *corpus* de pesquisa, utilizou-se a base de dados Scielo e os seguintes termos foram utilizados para a busca: “terminologia”, “exaustividade” e “especificidade”. Esses termos foram buscados de forma combinada, por meio da busca avançada, nas áreas de maior conteúdo informacional: título, resumo e palavras-chave e não houve delimitação temporal. Foram pré-definidos os periódicos Ciência da Informação, Perspectivas em Ciência da Informação e Transinformação, por serem periódicos reconhecidos da área e que trariam os termos dentro do universo de pesquisa da Ciência da Informação. Com esses termos de busca, não foram recuperados nenhum periódico que tratassem exclusivamente sobre indexação ou sobre terminologia. O corpus totalizou dezesseis artigos, sendo que sete possuíam conceitos de terminologia que poderiam ser utilizados na pesquisa.

Posteriormente, foi elaborado um quadro com as definições dos três termos, de forma comparativa.

6 Resultados

Abaixo, têm-se o quadro comparativo, demonstrando que conceitos da terminologia podem ter correspondência com os conceitos de exaustividade e especificidade.

Pode-se perceber que as duas áreas – indexação e terminologia - apresentam convergência entre os conceitos, identificando que é possível um maior diálogo entre elas. Considerando que os dois campos tem por preocupação a elaboração de termos representativos, possuindo, portanto, o mesmo objetivo, características como a exaustividade e a especificidade devem estar presentes no momento dessa elaboração.

Quadro 2 – Comparação entre os conceitos de terminologia, exaustividade e especificidade.

TERMINOLOGIA	EXAUSTIVIDADE	ESPECIFICIDADE
"Na terminologia, a ênfase é dada ao desenvolvimento de glossários, que são instrumentos de padronização de termos de uma dada especialidade ou domínio." (SALES; CAMPOS; GOMES, 2008, p. 66)	Tratando de um domínio, deve então ser exaustivo.	Termos referentes a uma especialidade devem ser tão específicos quanto a dada especialidade.
"(A terminologia é) qualquer atividade relacionada com a sistematização e representação de conceitos ou apresentação de termos baseados em princípios e métodos estabelecidos e ainda como um conjunto de termos que constituem um sistema de conceitos de uma determinada área" (DIAS, 2000, p. 90)	A exaustividade diz respeito ao número de conceitos que serão estabelecidos aos termos, constituindo, portanto, um sistema de conceitos.	
"(...) como produto, é um conjunto de termos, ou vocabulário, de uma determinada especialidade." (DIAS, 2000, p. 90)	A exaustividade diz respeito ao número de termos que serão representativos de dado documento.	Por se tratar de uma especialidade, é portanto, específico.
"Se tomada como objeto, isto é, a terminologia como um conjunto de termos de uma		Tratando-se de uma especialidade os termos

especialidade, nota-se que cada área do conhecimento aborda seus termos de forma diferente." (DIAS, 2000, p. 90)		devem então, ser específicos.
"Uma terminologia especializada traz subsídios ao indexador para representar com maior precisão os documentos que estarão disponíveis em bases de dados para a comunidade científica realizar suas buscas." (MARTINS; CARVALHO, 2014, p. 122)	A exaustividade é, da mesma forma, uma atividade que subsidia o trabalho do indexador.	A especificidade é, da mesma forma, uma atividade que subsidia o trabalho do indexador, além de estar diretamente relacionada à precisão da representação.
"A terminologia, como meio de comunicação formal entre especialistas de um domínio, é constituída por termos monorreferenciais: cada termo tem um conceito específico e se relaciona com os demais por intermédio de sua definição, permitindo a precisão e normalização desta linguagem." (ALCAIDE et al., 2001, p. 26)	Por representar um domínio, deve, portanto, ser exaustiva.	Cada termo possuir um termo específico exige especificidade. Além do que, um dos princípios da especificidade é justamente, a precisão da recuperação.
"o termo é uma unidade de comunicação que representa o conceito, seu significado é denotativo, composto por palavras determinadas por especialistas de uma área específica do conhecimento." (BORBA; VAN DER LAAN; CHINI, 2012, p. 32)		Sendo assim, o termo representativo de uma área específica do conhecimento, exige especificidade.
"(...) o conhecimento do assunto (isto é, dos conceitos de determinada área) é elemento norteador para a criação do termo." (BORBA; VAN DER LAAN; CHINI, 2012, p. 32)	Para conhecer o assunto e os conceitos de determinada área, deve-se assim, ser exaustivo.	
"A terminologia empregada na avaliação científica pode ser enquadrada entre os indicadores qualitativos, porque entre seus pontos fortes está sua relação com		Por se tratar de conteúdos específicos, exige especificidade.

os conteúdos específicos das ideias científicas representadas pelos termos." (MORAES, 2007, p. 33)		
"A Teoria Geral da Terminologia busca a objetividade e a precisão, com base na filosofia analítica." (MORAES, 2007, p. 34)		A especificidade está diretamente ligada à precisão dos termos.
"Para sanar esse problema, iniciou-se uma padronização operacional terminológica sem, no entanto, deixar de respeitar a especificidade do termo" (MORAES, 2007, p. 35)		Novamente, o autor remete à noção de especificidade.
"Houve casos de termos que (...) apareciam com muita especificidade. No entanto, os pesquisadores da Fiocruz optaram por um termo mais abrangente." (MORAES, 2007, p. 36)	Em determinados domínios do conhecimento, em alguns momentos os termos devem ser mais abrangentes e, portanto, mais exaustivos.	
"Porém, mesmo quando não se trata de universos especializados, ou seja, trata-se de universos cujos contornos não são bem determinados, os princípios terminológicos associados aos documentários, são essenciais para referir o processo de organização espacial e visual dos elementos do repertório em jogo" (LARA, 2004, p. 243)	Novamente, alguns domínios do conhecimento não são delimitados, sendo, portanto, necessária uma terminologia mais abrangente e exaustiva.	

Fonte: os autores.

7 Considerações finais

O presente trabalho buscou demonstrar como os conceitos de terminologia podem se correlacionar com os conceitos de exaustividade e especificidade, encontrados na indexação. Essa busca deu-se a partir da literatura brasileira da área de Ciência da

Informação e resultou em um quadro comparativo, no qual trechos da literatura foram extraídos, quando essas correlações eram encontradas.

Percebeu-se que alguns conceitos de terminologia podem, portanto, se relacionar com os conceitos básicos de exaustividade e especificidade na indexação. Principalmente considerando que uma terminologia deve ser exaustiva para compreender todo um domínio do conhecimento e também específica, para representar bem as especialidades de dado conhecimento.

Essa pesquisa, tendo caráter exploratório, procurou demonstrar apenas que existem essas evidências e, portanto, sugere-se que seja dada continuidade, de forma a procurar esses conceitos nas obras dos principais teóricos da área.

Apoio: Fundação de Amparo à Pesquisa do Estado de São Paulo.

Referências

ALCAIDE, G. S. et al. Análise comparativa e de consistência entre representação automática e manual de informações documentárias. **Transinformação**, v. 13, n. 1, p. 23-41, 2006.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676:** Métodos para análise de documentos - Determinação de seus assuntos e seleção de termos de indexação - Procedimento. Rio de Janeiro: ABNT, 1992.

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (org.). **Educação, Universidade e Pesquisa**. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p.35-60.

BORBA, D. S.; VAN DER LAAN, R. H.; CHINI, B. R. Palavras-chave: convergências e diferenciações entre a linguagem natural e a terminologia. **Perspectivas em Ciência da Informação**, v. 17,n. 2, p. 26-36, 2012.

CABRÉ, M. T. **La terminología: teoría, metodología, aplicaciones**. Barcelona: Antártida/Empúries, 1993. 529p

CERVANTES, B. M. N.; RUBI, M. P.; FUJITA, M. S. L. Terminologias em política de indexação**. Ibersid (Zaragoza)**, v. 2, p. 211-222, 2008.

COLLISON, R. L. **Índices e indexação: guia para indexação de livros e coleções de livros, periódicos, partituras musicais, discos, filmes e outros materiais.** Tradução por Antônio Agenor Birquet de Lemos. São Paulo: Polígono, 1972, 223 p.

DIAS, C. A. Terminologia: conceitos e aplicações. **Ciência da Informação**, v. 29, n. 1, p. 90-92, 2000.

FAULSTICH, E. Aspectos de terminologia geral e terminologia variacionista. **TRADTERM**, v. 7, São Paulo, 2001.

FUJITA, M. S. L.; RUBI, M. P. Elementos de política de indexação em manuais de indexação de sistemas de informação especializados. **Perspectivas em Ciência da Informação**, v. 8, n.8, p. 38-54, 2003.

FUJITA, M. S. L. (Org). **A indexação de livros a partir da percepção de catalogadores e usuários de bibliotecas universitárias: um estudo de observação do contexto sociocognitivo com protocolos verbais**. São Paulo: Cultura Acadêmica, 2009.

HJøRLAND, B. What is knowledge organization?. **Knowledge Organization**, v. 35, n. 3/2, 86-101, 2008.

KRIEGER, M. G. Terminologia revisitada**. D.E.L.T.A.,** v. 16, n. 2, 209-228, 2000.

LANCASTER, F. W. **Indexação e resumos: teoria e prática.** Tradução de Antonio Agenor Briquet de Lemos. Brasília: Briquet de Lemos, 2004.

LARA, M. L. G. Linguagem documentária e terminologia. **Transinformação**, v. 16, n. 3, p. 231-240, 2004.

MARTINS, E. C. D.; CARVALHO,T. Recuperação da informação em psicologia: LILACS e Index Psi Revistas Tecno-Científicas. **Perspectivas em Ciência da Informação**, v. 19, n. 2, p. 118-130, 2014.

MATURANA, M. T. I. **Analysis y recuperación de información documental.** Santiago: Ediciones Instituto Profesional de Santiago, 1992, 237 p.

MORAES, A. F. Terminologia como indicador qualitativo. **Transinformação**, v. 19, n. 1, p. 31-38, 2007.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **ANSI/NISO Z39.19:** guidelines for the construction, format, and management of monolingual controlled vocabularies. Estados Unidos: NISO, 2010.

NATIONAL INFORMATION STANDARDS ORGANIZATION.**NISO TR02:** guidelines for indexes and related information retrieval devices. Estados Unidos: NISO, 1997.

SALES, L. F.; CAMPOS, M. L. A.; GOMES, H. E. Ontologias de domínio: um estudo das relações conceituais. **Perspectivas em Ciência da Informação**, v. 13, n. 2, p. 62-76, 2008.

Aspectos dos metadados de assunto na catalogação de fotografias

Ana Carolina Simionato
Universidade Federal de São Carlos
acsimionato@ufscar.br

1 Introdução

O surgimento de novas possibilidades para a localização de imagens em meio digital, revelou que algumas formas de representação ainda permanecem com falhas que prejudicam a sua busca e acesso. Seja no ambiente tradicional ou digital, esse fato ainda carece de discussões, em razão das grandes dificuldades na definição dos metadados, descritivos e assunto para as fotografias.

Em suas particularidades as fotografias devem ser consideradas como ponto primordial, ou seja, o recurso informacional deve ser a frente de qualquer forma de representação. Preferencialmente não sendo atreladas a uma regra geral que concentra diversos tipos de recursos informacionais, a exemplo de regras gerais, mas sim, conceituando a própria solução do tipo imagético.

O trabalho então possui o objetivo de verificar os aspectos da análise de assunto na representação de fotografias, no intuito de apresentar formas e instrumentos que auxiliam na definição do valor para o campo, o metadado de assunto, incluindo as características dos sistemas de organização do conhecimento e da informação.

Nesse sentido, não será escopo desse trabalho discutir sobre as perspectivas teóricas de desenvolvimento de cada instrumento para representação das fotografias, mas sim, apontá-los como possibilidades de forma eficiente para localização, recuperação, acesso e uso. Dessa forma, a metodologia utilizada é de natureza teórica,

exploratória e qualitativa em relação a dinâmica ao mundo e ao sujeito, em razão da interpretação e compreensão dos fenômenos. (GIL, 2002).

O trabalho corresponde a uma parte concluída da pesquisa em andamento sobre o tratamento temático e descritivo de acervos fotográficos do estado do Panará - Brasil, desenvolvido pelo grupo de pesquisa Organização e Representação da Informação e do Conhecimento de Recursos Imagéticos - ORICRI. Por meio do levantamento bibliográfico realizado, foi encontrado diversas formas de realizar a atribuição de valor do metadado assunto, no entanto, não foi encontrado orientações conjuntas para que isso seja feito sistematicamente ao que é encontrado à Ciência da Informação ou mesmo, eram desconhecidas pelas instituições que representavam as fotografias.

Os aspectos de metadados de assunto na catalogação de fotografia, foi dividido em tópicos que designam ao que se referem a representação das fotografias, as possibilidades para o metadado de assunto e as considerações finais do trabalho.

2 A Representação das Fotografias

A partir da invenção da câmera escura no século XIX e a garantia de fixação em uma placa para processos comerciais, em 1839 conhecida como daguerreotipo, o registro fotográfico se popularizou com câmeras portáteis de custo reduzido que hoje foram transpostas ao meio digital.

Consequentemente, com o crescimento do número de câmeras a proporção de registros fotográficos foi exponencial, dessa forma, tomaram proporções em que possibilitaram o reconhecimento de seu valor histórico e cultural, justificando o armazenamento nas unidades de informação. Por meio do armazenamento dessa nova possibilidade de acervo, surgem dificuldades na representação desses recursos informacionais.

As câmeras digitais e suas imagens lidam fluentemente com os metadados, atributos que caracterizam um recurso informacional, no caso os recursos imagéticos. É possível dizer que os metadados são as formas mais simples de descrição em ambientes informacionais digitais, são compostos por atributos e valores, previamente definidos por uma forma metodológica para composição de um padrão de metadados.

As imagens digitais possuem automaticamente metadados que a descrevem de forma técnica, denominados como dados *Exchangeable Image File Format* (EXIF), esses metadados são informações gravadas conjunto ao registro em relação a 3.5 mm de distância focal, por exemplo. Entretanto, nas fotografias tradicionais, os profissionais devem escolher a melhor forma para descrição das informações técnicas e no caso, independente do meio, ambas devem elucidar a representação por assunto.

A representação, pode ser conceituada como um conjunto de convenções sintáticas e semânticas que tornam possível descrever as coisas, de uma forma intrínseca e extrínseca, e necessita de padrões para formalizar e adaptar a apresentação prevista e proposta ao usuário, no sentido de aperfeiçoar suas buscas, acesso e localização. Nesse sentido, Smit (1996, p. 29) aponta que

> A proposição de uma metodologia de análise da fotografia supõe um entendimento da essência desta, daquilo que a caracteriza, das razões pelas quais e produzida e, sobretudo, das condições em que será utilizada. Em outras palavras, torna-se necessário compreender a imagem fotográfica, enquanto informação a ser tratada e recuperada.

Entretanto, a padronização almejada pelos bibliotecários, não é a garantia de uma representação adequada para uma fotografia e os instrumentos de representação ainda não oferecem amparo eficiente. (SIMIONATO, 2012).

Os códigos tradicionais da área de Biblioteconomia não conseguem abarcar a especificidade da imagem. A exemplo, o Código de Catalogação Anglo-Americano (AACR2r) que não apresenta subsídios necessários para descrição das técnicas fotográficas e do próprio assunto principal. Com isso, o profissional deve estar sempre consultando outros padrões de metadados, como o CDWA, VRA *Core*, *Cataloging Cultural Objects*, *Graphic Materials* ou mesmo o catálogo decisório da instituição.

Caso a instituição utilize o MARC 21, os campos e subcampos disponíveis serão encontrados metadados de assunto no grupo 6XX, correspondendo ao 650 para termo tópico e o 690 para o termo atribuído à instituição.

Nesse contexto, o que abre para discussão e o para o tema desse trabalho, é que as estruturas atualmente estabelecidas correspondem em padrões e metodologias aceitas e facilitadas aos bibliotecários, no entanto, as informações correspondentes ao

valor dos metadados, que em muitos casos dificulta a compreensão e ainda, sem a devida padronização de instruções para o preenchimento, afetam a interoperabilidade.

3 As possibilidades para os Metadados de Assunto

A representação de uma imagem para os fotógrafos é denominada como *metalogging*, que consiste na inclusão de informações de forma que possa ser armazenada e utilizada como uma ajuda na recuperação de que a imagem a partir de um banco de dados ou coleção depois de ser registrado. (DALY, 2014). Sendo que grande parte da dificuldade do preenchimento dos metadados está no campo assunto, o qual os fotógrafos denominam como processo de *keywording*, atribuindo termos que aumentam a visibilidade e permitem pesquisas mais relevantes. (DALY, 2014).

Este campo depende das possibilidades que o ambiente ou mesmo a instituição oferece, por essa razão, muitas formas de representação por assunto são totalmente livres ao usuário, assim denominada a *folksonomia*.

Entre alguns ambientes digitais como *Flickr*, *Picasa*, *Zooomr*, Pixlr, entre outros, elegeram a *folksonomia* como a principal forma de recuperação atribuída aos usuários a responsabilidade dos termos, através de termos da linguagem natural. Dessa forma, os usuários categorizam e indexam temas que julgam estarem relacionados a imagem, contudo na maioria das vezes estas *tags* acabam representando um conteúdo que não é o que foi apresentado. Como Herst et al. (2007, não paginado, tradução nossa) apontam que “[...] um computador procura as imagens na Internet através de palavras-chave ou *tags*, porém nem sempre é encontrado o que se deseja; isso acontece, pois, as imagens são encontradas com base nos metadados que as acompanham [...]”.

Ao pensar que a *folksonomia* e *keywording* em muitos casos não conseguem garantir a melhor recuperação da imagem, os profissionais da informação apresentam alguns instrumentos da que podem ser parte desse processo de definição de assunto.

Vale apontar que a necessidade de criação de atributos ou mesmo a atribuição adequada de valores representativos a estas informações, para que apresentem adequadamente aos usuários, por meio de uma recuperação eficiente e eficaz. Ou seja,

os instrumentos para padronização dessas informações evitam alguns equívocos como usar ‘branco e preto’, um valor descritivo em um valor temático.

Para isso, é necessário fazer uma análise e síntese do conteúdo imagético e posteriormente realizar a descrição e os metadados, esses procedimentos estão inseridos as principais operações da análise documental, sendo ainda incluídos os processos de classificação, indexação e condensação documentária.

Segundo as autoras Boccato e Fujita (2006, p. 90)

> A análise documental de imagens deve atender aos preceitos da documentação, refletindo a credibilidade e segurança no momento da recuperação da informação pelo usuário. Atendendo assim ao objetivo central da análise documental, a informação documental deve promover a identificação de materiais informacionais que respondam, de maneira satisfatória, às questões dos usuários e, por outro lado, possibilitar a tomada de decisões sobre a consulta e a escolha de um determinado documento original.

A análise documental deve-se pautar principalmente em uma análise do conotativo concreto (RODRIGUES, 2011) para interpretação da fotografia deve ser observada pelos os elementos constitutivos da mesma, com termos que sejam claros, exemplo: arquitetura, prefeitura municipal, cidade.

Essa análise por meio do conotativo concreto diminui a subjetividade da análise e descrição do conteúdo. Erwin Panofsky (1979) estabeleceu níveis para análise da imagem que orienta nesta subjetividade. Panofsky (1979) apresenta três níveis para a análise de imagem: 1) pré-iconográfico, onde são descritos genericamente os objetos e ações representadas pela imagem; 2) iconográfico, estabelece o assunto secundário ou convencional ilustrado pela imagem. Ou seja, trata-se da determinação do significado abstrato ou simbólico da imagem, sintetizado a partir de seus elementos componentes, detectados pela análise pré-iconográfica; e 3) iconológico que propõe uma interpretação do significado intrínseco do conteúdo da imagem. A análise iconológica é construída a partir dos níveis anteriores, mas possui influências do conhecimento do analista sobre o ambiente cultural, artístico e social no qual a imagem foi gerada.

A imagem é, simultaneamente, específica e genérica, segundo Shatford (1986) baseado nos níveis de Panofsky (1979) discute a representação da imagem, introduzindo

uma distinção entre de genérico, de específico e sobre. Shatford (1986) representa os níveis em categorias, pelo motivo que um usuário pode compreender o sentido pré-iconográfico de uma imagem (por exemplo: rio), mas ele não consegue entender suas necessidades em termos iconográficos (por exemplo: Rio Amazonas).

As categorias de Shatford (1986) representam entre os questionamentos de: Quem? (seres), Onde? (espaço), Quando? (tempo), Como? (técnica) e O que? (ação). Estas categorias também foram aplicadas e estudadas por outros autores, que citamos: Smit (1989), Kossoy (2001), Manini (2001) e Maimone e Gracioso (2007).

No trabalho de Costa (2008) estas categorias também foram relacionadas com a categorias Personalidade, Matéria, Energia, Espaço e Tempo - PMEST, ditas como 'categorias essenciais', propostas por Ranganathan em 1931. Assim, encontrando relação com Personalidade - Quem?, Matéria - O que?, Energia - Como?, Espaço - Onde? e por fim, Tempo - Quando?.

Por conseguinte, a análise documental para atribuição do valor ao metadado de assunto, existem outros processos vinculados que envolvem além das categorias, o processo é a indexação. A partir da indexação, processo originário da necessidade de um aprofundamento de conteúdos específicos, é possível identificar o tema de um documento e representar por conceitos, ou seja, transformar a linguagem natural - LN para uma linguagem documentária - LD. (FUJITA, 2009).

A partir desse processo terminológico são criados instrumentos para padronização de termos, esses instrumentos variam em seu nível de aprofundamento ao conteúdo e bem como a sua forma de padronização terminológica.

Entre os instrumentos, os vocabulários controlados se destacam por serem sistemas de organização do conhecimento (BOCCATO; RUBI; FUJITA, 2014), e que possibilitam a representação dos conceitos identificados e selecionados mediante a análise conotativa concreta do conteúdo fotográfico.

Segundo Harpring (2010, p. 166, tradução nossa) o vocabulário controlado na "[...] catalogação é uma maneira mais eficiente de incorporar terminologias, sendo esta por meio da utilização de autoridades locais ou incluindo os vocabulários controlados publicados em sua totalidade.".

Dessa forma, os vocabulários controlados estão dispostos em uma ordem específica e caracterizam por três relações: equivalência, hierárquica e associativa. Quando seus termos são estruturados de forma hierárquica, condizem a formação de um tesauro, que pode ser tanto monolíngue quanto multilíngue, e os objetivos concentram em promover a coerência na indexação de conteúdo e para facilitar a pesquisa e navegação. (HARPRING, 2010).

Os vocabulários controlados para recursos imagéticos possuem estrutura e repertório terminológico que possibilitam a representação e análise para a inserção no metadado de assunto, no caso, será exemplificado pela iniciativa do Instituto *Getty Research Institute*, dos Estados Unidos.

O *Getty Research Institute* construiu o conjunto do *Art & Architecture Thesaurus* (AAT), a *Getty Thesaurus of Geographic Names* (TGN) e em construção a *The Cultural Objects Name Authority* (ULAN) todos são vocabulários estruturados que podem ser usados para melhorar o acesso à informação sobre arte, arquitetura, e cultura material. Os termos dos tesauros e vocabulários controlados do *Getty Research Institute* são recolhidas ou são relacionados a partir de coleções de museus, coleções especiais, arquivos, bibliotecas, pesquisas acadêmicas e outras fontes. Os tesauros podem ser acessados pelo link: <http://www.getty.edu/research/tools/vocabularies/>.

Outra característica importante é que eles estão sendo construídos para permitir seu uso em dados vinculados, trabalhando conjuntamente com a iniciativa *Linked Open Data*, tornando parte do esforço contínuo do *Getty Research Institute* para que os recursos de conhecimento sejam disponíveis gratuitamente a todos.

4 Considerações finais

Entre os apontamentos já realizados durante o trabalho, elucida que a análise documental, os vocabulários controlados e os tesauros, aumentam a consistência metadados nos sistemas de recuperação para a localização, acesso, uso e reuso de fotografias, em bancos de dados, catálogos e outras formas de armazenamento imagético.

Nesse sentido, é importante que os profissionais da informação discutam seriamente a necessidade da utilização de formas mais estruturadas para o preenchimento dos metadados de assunto, bem como discutam com seus principais usuários, os fotógrafos e demais usuários especializados, uma transformação dos termos naturalmente utilizados para que o sistema seja melhor e reflita a interlocução entre ambos.

Portanto, a utilização de estruturas terminológicas, hierárquicas ou não, fornecem ferramentas importantes para recuperação de recursos imagéticos, seja ele em diferentes lugares, em diferentes idiomas e isso poderá acarretar em novas propostas e cada vez mais tecnológicas, para a vinculação e relacionamento de registros, conjunto a proposta do *Linked Data*.

Referências

BOCCATO, V. R. C.; FUJITA, M. S. L. Discutindo a análise documental de fotografias: uma síntese bibliográfica. **Cadernos BAD**, n. 2, 2006, p. 84-100.

BOCCATO, V. R. C.; RUBI, M. P.; FUJITA, M. S. L. Aplicabilidade de vocabulário controlado na análise documental de fotografias: uma prática possível. In: SOUZA, L. M. A.; FUJITA, M. S. L.; GRACIOSO, L. (Org.). **A imagem em Ciência da Informação**: reflexões teóricas e experiências práticas. Marília: Oficina Universitária; Cultura Acadêmica, 2014, p. 117-152.

COSTA, L. S. F. **Uma contribuição da teoria literária para a análise de conteúdo de imagens publicitárias do fim do século XIX e primeira metade do século XX, contemplando aspectos da natureza brasileira**. 263f. 2008. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Filosofia e Ciências, 2008.

DALY, T. **The fundamentals of digital photography**. Londres: A&C Black, 2014.

FUJITA, M. S. L. (Org.) **A indexação de livros**: a percepção de catalogadores e usuários de bibliotecas universitárias. São Paulo: Cultura Acadêmica, 2009. Disponível em: http://www.culturaacademica.com.br/catalogo-detalhe.asp?ctl_id=56. Acesso em: 20 mai. 2015.

GIL, A. C. **Metodologia da pesquisa**. São Paulo: Atlas, 2002.

HARPRING, P. **Introduction to controlled vocabularies**: terminology for art, architecture, and other cultural works. Los Angeles: Getty Publications, 2010.

HEARST, M. et al. **Faceted metadada for image search and browsing**. Berkeley: University of California, 2007.

KOSSOY, B. **Fotografia & história**. São Paulo: Ateliê, 2001.

MAIMONE, G. D.; GRACIOSO, L. S. Representação temática de imagens: perspectivas metodológicas. **Informação & Informação**, v. 12, n. 1, 2007. Disponível em: http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/1760/1504. Acesso em: 20 mai. 2015.

MANINI, M. P. Análise documentária de imagens. **Informação & Sociedade**, v. 11, n. 1, 2001. Disponível em: http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/313/236. Acesso em: 20 mai. 2015.

PANOFSKY, E. **Significado nas artes visuais**. 2.ed. São Paulo: Perspectiva, 1979.

RODRIGUES, R. C. **Análise e tematização da imagem fotográfica**: determinação, delimitação e direcionamento dos discursos da imagem fotográfica. 2011. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, 2011.

SHATFORD, S. Analyzing the subject of a picture: a theoretical approach. **Cataloging & Classification**, v. 6, n. 3, p. 39-62, 1986.

SIMIONATO, A. C. **Representação, acesso, uso e reuso da imagem digital**. 142f. 2012. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista “Júlio de Mesquita Filho”, Faculdade de Filosofia e Ciências, Marília/SP, 2012.

SMIT, J. W. A análise da imagem: um primeiro plano. In: SMIT, J. W. (Org.) **Análise documentária**: a análise da síntese. Brasília: IBICT, 1989.

SMIT, J. W. A representação da imagem. **Informare**: cadernos do Programa de Pós-graduação em Ciência da Informação), v.2, p. 28 -36, 1996.

Linguagens documentárias e tratamento de recursos audiovisuais: interfaces teóricas e metodólogicas

Francisnaira Cristina Ravazi
Diretoria Regional de Ensino de Tupã
francisnaira@gmail.com

Walter Moreira
Universidade Estadual Paulista
walter.moreira@marilia.unesp.br

1 Introdução

Ao longo da história, a biblioteconomia e a documentação tem se preocupado com o desenvolvimento de instrumentos de organização e representação da informação registrada em documentos impressos. Os recursos audiovisuais receberam, por razões diversas, menor atenção. Nesse sentido, não se encontram instrumentos de classificação, por exemplo, que considerem as especificidades de um cenário que envolve a "cultura audiovisual eletrônica". Entre outras coisas, a oferta de outros modos de ler e perceber questionam a centralidade da relação escola-livro como núcleo do sistema escolar vigente (MARTÍN-BARBERO; REY, 1999).

É possível inferir que esse modelo (relativamente) novo de cultura, proporciona ao usuário da informação, especialmente aos mais jovens, novos modos de leitura que interferem, naturalmente, em sua visão de mundo e no modo como irá *consumir* a informação em seus diferentes níveis.

O avanço realizado nas teorias e metodologias desenvolvidas pela biblioteconomia e pela documentação para a organização do conhecimento, considerando-se a perspectiva de seu uso social, é limitado em relação às especificidades da linguagem audiovisual. Nos campos da comunicação e da pedagogia, há alguns avanços teóricos

consideráveis, embora não haja nesse caso preocupações com o que se pode denominar *organização profissional dos recursos audiovisuais*.

Dada a importância e a popularidade dos recursos audiovisuais, é natural que haja antecipações de soluções, pelos mais diversos setores interessados na questão, em relação às áreas mais clássicas preocupadas com a organização, a representação e a recuperação da informação. É necessário e extremamente útil à ciência da informação, entretanto, identificar, descrever e analisar os modelos utilizados – profissionalmente ou não – para a representação documentária dos recursos audiovisuais.

Propôs-se como objetivo geral para esse trabalho, identificar a zona de confluência teórica e metodológica entre as linguagens documentárias e o tratamento dos recursos audiovisuais. O objetivo geral foi complementado operacionalmente com os seguintes objetivos específicos: a) discutir o conceito de audiovisual e b) identificar e comparar, por meio de sistematização, as metodologias de organização da informação audiovisual.

Para a realização da pesquisa foram utilizados dois tipos de *corpora*: a) *corpus* documental: do qual foram extraídos os artigos e trabalhos para análise; b) *corpus* de análise: os artigos e trabalhos selecionados do *corpus* documental mediante estratégias de busca afinadas com os objetivos desta pesquisa.

O *corpus* documental foi composto por todo o conjunto de periódicos disponível na Base de Dados Referencial de Artigos de Periódicos de Ciência da Informação (BRAPCI) e nos anais do ENANCIB. Em ambos os casos, observou-se um recorte cronológico de 13 anos, no período 2000-2012.

Na realização das buscas para a composição do *corpus* de análise, foram utilizadas estratégias que contemplavam os seguintes termos: “classificação”, “linguagem de indexação”, “linguagem documental”, “linguagem documentária”, “organização da informação”, “audiovisual”, “recurso audiovisual”, “organização do conhecimento”, “representação da informação”, “representação do conhecimento”, “tesauro”, “ontologia”. Observou-se de modo preferencial a presença dos termos nos campos de palavras-chave e títulos dos trabalhos e, de modo complementar, nos resumos. Os termos foram propostos de forma isolada e com qualificadores, considerando-se também suas variações de número.

Para a análise, os trabalhos localizados foram categorizados por temática, sendo: a) linguagens documentárias em geral; b) sistemas de classificação; c) tesauros; d) ontologias; e) audiovisual. Como interessam à pesquisa de modo particular os trabalhos que contemplam teorias e/ou metodologias sobre organização da informação audiovisual, foram formadas mais quatro categorias decorrentes do entrecruzamento das anteriores, quais sejam: f) linguagens documentárias e audiovisual; g) sistemas de classificação e audiovisual; c) tesauros e audiovisual; d) ontologias e audiovisual.

Categorizados os artigos, procedeu-se o desenvolvimento da análise, com a finalidade de identificar conceitos, aportes teóricos e metodologias aplicadas na solução de problemas relativos à organização e representação da informação audiovisual.

2 Conceito de audiovisual

O conceito de audiovisual ainda não é estável e sua identificação requereu uma análise cuidadosa da literatura, dada a sua centralidade nessa pesquisa.

Utilizando-se a expressão de busca “audiovisual” na BRAPCI, foram identificados 27 artigos na temática do trabalho. Na utilização dos outros termos de busca propostos, (“audio visual”, “audiovisuais” e “audio visuais”), foram encontrados menos artigos, mas iguais aos que já haviam sido levantados com o termo “audiovisual”, atestando já de início a oscilação terminológica que ainda existe sobre o tema conforme observaram diversos autores, entre esses, Bethônico (2006). Essa oscilação na disponibilização da informação prejudica o usuário no momento de recuperação da informação e dificulta, evidentemente, a pesquisa bibliográfica sobre o assunto, além de comprometer a precisão do conceito.

O Dicionário de Biblioteconomia e Arquivologia define **material audiovisual** como um “suporte de informação que não pode prescindir de equipamento para audição ou visão, que revele o seu conteúdo” (CUNHA; CAVALCANTI, 2008, p. 241). O Dicionário Brasileiro de Terminologia Arquivística classifica **documentação audiovisual** como “gênero documental integrado por documentos que contêm imagens, fixas ou em movimento, e registros sonoros, como filmes e fitas videomagnéticas” (ARQUIVO NACIONAL, 2005, p. 73). Ainda no Dicionário de Biblioteconomia e Arquivologia, o termo

documento audiovisual é definido como “documento que reproduz imagens fixas ou móveis, bem como registros sonoros em qualquer suporte, e que exige equipamento apropriado para ser visualizado ou executado” (CUNHA; CAVALCANTI, 2008, p. 133). Nos dois dicionários o **documento audiovisual** é descrito como um documento que apresenta som e imagem ao mesmo tempo, diferente, portanto, do que está descrito no termo **material audiovisual**.

A *International Encyclopedia*, editada por Feather e Sturges (2003, tradução livre), define materiais audiovisuais como um

termo genérico para descrever o conteúdo de informação disponibilizado em meios de armazenamento e transmissão e formatos que usam imagens e som, em vez de, ou, algumas vezes, em adição a, matéria textual. Inclui: CD de áudio, discos e fitas, fotografias, slides, filmes e vídeos e formatos que podem combinar dois ou mais destes [...]. O termo, no entanto, ainda é bastante útil, e não foi inteiramente substituído por multimídia, o qual é muitas vezes, de forma errada, usado como sinônimo da moda para ele.

Alguns autores (CEBRIAN-HERREROS, 1983; BETHÔNICO, 2006) defendem que o audiovisual deve apresentar, simultaneamente, imagem e som. Nessa linha, Bethônico (2006, p. 59) defende que o cinema mudo, por exemplo, “não é realmente audiovisual sem a música que participa da construção dos climas e que antigamente era tocada ao vivo nas sessões”.

Outros estudiosos questionam a simultaneidade de imagem e som, como Araújo (1992), Smit (1993) e Dieuzeide (1965). Este último define os audiovisuais como “meios mecânicos ou eletrônicos de registro, reprodução e difusão de mensagens sonoras ou visuais utilizados, separada ou conjuntamente, para apresentar conhecimentos, facilitar sua aquisição ou modificar determinados comportamentos”.

Contudo, no que caracteriza o audiovisual, a questão de apresentar som e imagem ao mesmo tempo é tratada por quase todos os autores que trabalham com este recurso, sem que se chegue a um consenso sobre a questão.

No que diz respeito ao alcance do conceito, compartilha-se neste trabalho, da concepção de que o audiovisual precisa apresentar som e imagem ao mesmo tempo. São

considerados audiovisuais, portanto, recursos como filmes, vídeos, videoclipes e outros, desde que apresentem como característica a presença simultânea de som e imagem.

3 Base bibliográfica dos recursos audiovisuais

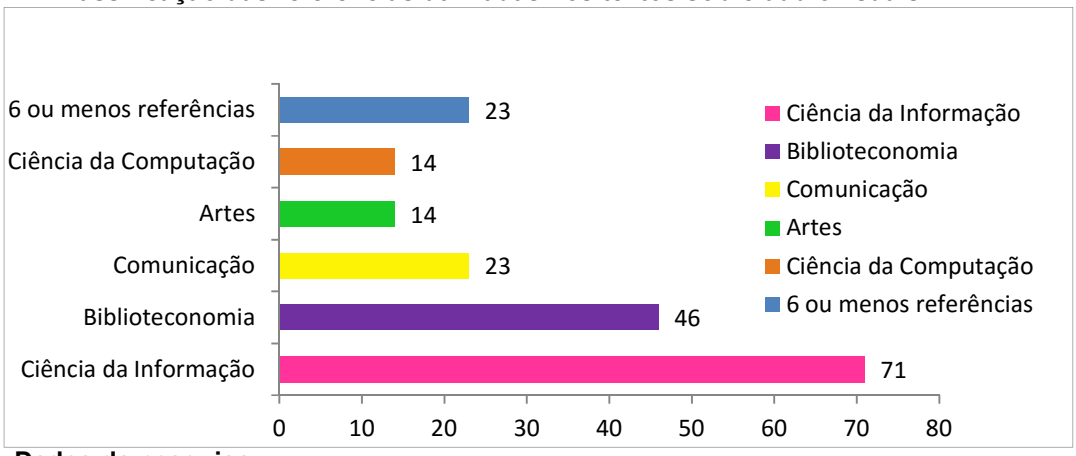
Para a análise da base bibliográfica, foram selecionados todos os textos recuperados no levantamento bibliográfico, na BRAPCI e no ENANCIB, utilizando-se o termo de busca “audiovisual”. Em cada artigo foram analisadas as referências bibliográficas com o fim de identiticar a área do conhecimento em que se situam.

Os autores/obras foram agrupados e ordenados pelo critério da quantidade de vezes em que foram referenciados. Desse modo, identificaram-se dois grandes grupos. O primeiro, com 18 autores, cujos trabalhos foram referenciados duas vezes. O segundo grupo com 173 autores/obras que receberam apenas uma referência.

Considerando-se que os textos analisados abordavam a temática do audiovisual, é expressivo que o número máximo de citações que um mesmo texto tenha recebido seja igual a 2, é igualmente expressiva a quantidade de autores que receberam apenas uma citação, isto é, foram mencionados apenas uma vez. Nesse caso é possível inferir que não há consenso sobre a literatura que embasa as pesquisas sobre audiovisual.

As referências utilizadas nos trabalhos selecionados também foram classificadas em áreas, conforme a tabela de áreas do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), com o objetivo de mapear as áreas de fundamentação para os estudos sobre recursos audiovisuais. O resultado está expresso na Figura 1. As áreas que receberam 6 ou menos referências foram agrupadas num mesmo item, com objetivo de melhorar a visualização.

Figura 1 - Classificação das referências utilizadas nos textos sobre audiovisuais



Fonte: Dados da pesquisa

A Ciência da Informação e a Biblioteconomia têm oferecido os maiores subsídios para o estudos sobre audiovisuais. Isso pode ser tomado como natural nesse trabalho, pois foram utilizados como dados os trabalhos publicados em periódicos e eventos científicos dessas áreas. Por outro lado, é preciso observar também que isso pode ser revelador de falta de diálogo. A questão dos audiovisuais tem sido amplamente discutida na Comunicação e também na Educação.

4 Instrumentos utilizados na organização dos recursos audiovisuais

Os resultados apresentados na sequência referem-se à sistematização do conjunto maior de dados utilizados em trabalho de pesquisa anterior (RAVAZI, 2014).

O Quadro 1 apresenta a tipologia de instrumentos de organização da informação utilizados para o tratamento da informação em recursos audiovisuais nos textos analisados. As referências bibliográficas utilizadas na primeira coluna do quadro foram abreviadas por questões de visualização e aparecem de modo completo na lista de referências ao final do trabalho.

Quadro 1 - Tipologia de instrumentos utilizados

REFERÊNCIA	INSTRUMENTO
FAGUNDES; PRADO (1999). Videoteca digital.	Banco de dados realizado com recurso do *Microsoft Office*
GONÇALVES (1992). Os novos paradigmas da imagem em movimento.	Tesauro

MOURA et al. (2005). Linguagens de indexação em contextos cinematográficos.	Tesauro
BARRETO (2007). Desafios e avanços na recuperação automática da informação audiovisual.	Metadado
MIRANDA; GARBELINI (2011). Tratamento técnico da documentação audiovisual na TV da Universidade Federal de Goiás.	Analise de acervo, decupagem, elaboração de etiquetas e conservação.
SOUSA; DUQUE (2012). Uso de ontologia para recuperação da informação disponibilizada em vídeos por meio de indexação multimodal.	Indexação multimodal e ontologia
MUCHERONI; RIBEIRO; PAIVA (2012). Descrição dinâmica para documentos multimodais em ambiente digital.	Ontologia para documentos multimodais
SILVA; SOUZA (2012). Vocabulários e metadados para descrição de recursos multimídia.	Vocabulários (incluindo ontologias) e metadados baseados em *linked data*
ALVARENGA; SOUZA (2012). A Universidade Federal de Minas Gerais no contexto do acesso aberto à informação científica.	Tags
SOUSA; DUQUE (2011). Recuperação da informação disponibilizada em vídeos: uma abordagem utilizando ontologia, processamento de linguagem natural e descritores multimodais.	Ontologia e Indexação multimodal

Fonte: Dados da pesquisa

Conforme se pode identificar no Quadro 1, alguns autores utilizam instrumentos relativamente tradicionais da biblioteconomia, tais como tesauros e os metadados, enquanto outros autores utilizam as ontologias e a indexação multimodal, ambos sendo, também relativamente, novos objetos de estudos no campo da ciência da informação. Observa-se também que são utilizados instrumentos ou técnicas que não se utilizam de instrumentos da biblioteconomia, como é o caso do estudos de Fagundes e Prado (1999) e Miranda e Garbelini (2011).

Apressadamente, poder-se-ia inferir que os autores que utilizaram outros instrumentos que não os tradicionalmente atribuídos à ciência da informação, não são pesquisadores de temas próprios da ciência da informação, ou que não utilizam seu referencial teórico. Observou-se, contudo, que todos os autores mencionados no Quadro 1 possuem relação direta e próxima com a ciência da informação, conforme foi verificado no campo de afiliação e no campo de atuação de cada texto estudado.

5 Considerações finais

Por meio da análise realizada em conformidade com os objetivos do trabalho, foi possível verificar o conceito e a base bibliográfica que sustentam as concepções de organização da informação audiovisual. Os procedimentos de organização desse tipo de recurso é realizado tanto por meio de instrumentos da ciência da informação (tesauros, metadados, indexação multimodal, ontologia e *tags*) quanto por outros meios (banco de dados utilizando editor de texto, decupagem, elaboração de etiquetas).

Verificou-se que o conceito de audiovisual ainda não está suficientemente claro. Ainda não há uma resposta precisa à identificação do conceito e nem mesmo sobre uma tipologia que descreva precisamente quais são esses materiais e quais suas especificidades. Observou-se que mesmo no nível sintático, o termo ainda não é absolutamente estável, como foi possível verificar pela presença de formas variadas de grafia em artigos de periódicos de uma área que, vale lembrar, prima pela organização da informação e pelo controle de vocabulário.

Há que se investigar para compreender melhor o motivo pelo qual nenhum pesquisador utilizou, nos textos analisados, sistemas de classificação para a organização dos recursos audiovisuais. Talvez isso se deva ao fato dos sistemas de classificação não terem sido projetados especificamente para recursos digitais, conforme são compreendidos os recursos audiovisuais neste cenário, porém é fato que os sistemas de classificação são indispensáveis já que oferecem um lastro teórico e metodológico para soluções de problemas referentes à organização do conhecimento e ao tratamento temático da informação.

A proposta geral dessa pesquisa foi identificar quais são e como são aplicados os instrumentos de organização da informação na organização de recursos audiovisuais. Observou-se, contudo, que há ainda questões de fundo, como o próprio conceito de audiovisual que ainda requerem mais investigações. Assim, pretende-se em estudos futuros empreender, com o recurso da análise conceitual e terminológica essa tarefa.

Referências

ALVARENGA, L.; SOUZA, R. M. F. A Universidade Federal de Minas Gerais no contexto do acesso aberto à informação científica: identificação de seus sistemas de informação. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais**...2012.Rio de Janeiro, 2012. p. 21.

ARAÚJO, W. T. Uso da informação audiovisual em bibliotecas: dados de pesquisas. **Informação & Sociedade**, João Pessoa, v.2, n.1, p.35-41, jan./dez. 1992.

ARQUIVO NACIONAL (Brasil). **Dicionário Brasileiro de Terminologia Arquivística**. Rio de Janeiro: Arquivo Nacional, 2005.

BARRETO, J. S. Desafios e avanços na recuperação automática da informação audiovisual. **Ciência da Informação**, Brasília, v. 36, n. 3, p. 17-28, set./dez. 2007.

BETHÔNICO, J. Signos audiovisuais e Ciência da Informação: uma avaliação. **Encontros Bibli**, Florianópolis, 2° número esp., 2º sem. 2006.

CEBRIAN-HERREROS, M. Fundamentos de la teoria y técnica de la información audiovisual. Madrid: Mezquita, 1983. 2 v.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de Biblioteconomia e Arquivologia**. Brasília: Briquet de Lemos, 2008.

DIEUZEIDE, H. **Les tecniques audiovisuelles dans l'enseignement**. Paris: Puf, 1965.
FAGUNDES, M. L. F.; PRADO, G. S. Videoteca digital: a experiência da videoteca multimeios do IA/UNICAMP. **Transinformação**, v. 11, n. 3, p. 293-299, set./dez. 1999.

FEATHER, J.; STURGES, P. (Eds.). **International Encyclopedia of Information and Library Science**. 2. ed. London: Routledge, 2003.

GONÇALVES, A. C. B. Os novos paradigmas da imagem em movimento. **DataGramaZero**, v. 3, n.1, fev. 2002.
MARTÍN-BARBERO, J.; REY, G. **Los ejercicios del ver**: hegemonia audiovisual y ficción televisiva. Barcelona: Gedisa, 1999.

MIRANDA, M. L. C.; GARBELINI, M. F. Tratamento técnico da documentação audiovisual na TV da Universidade Federal de Goiás. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília. **Anais**...2011.Brasília, 2011. p. 439-451.

MOURA, M. A. et al. Linguagens de indexação em contextos cinematográficos: a experiência de elaboração do tesauro eletrônico do cinema brasileiro. **Perspec. Ci. Inf.,** v.10, n.1, p. 54-69, jan./jun. 2005.

MUCHERONI, M. L.; RIBEIRO, C.; PAIVA, D. C. Descrição dinâmica para documentos multimodais em ambiente digital. In: ENCONTRO NACIONAL DE PESQUISA EM

CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais**...2012.Rio de Janeiro, 2012. p. 17.

RAVAZI, F. C. **A organização dos recursos audiovisuais por meio das linguagens documentárias e das ontologias:** aspectos teóricos e metodológicos. 2014. 83 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, Marília/SP, 2014.

SILVA, D. L.; SOUZA, R. R. Vocabulários e metadados para descrição de recursos multimídia: uma perspectiva de organização e representação do conhecimento no espaço digital. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais**...2012.Rio de Janeiro, 2012. p. 7.

SOUSA, E. E.; DUQUE, C. G. Recuperação da informação disponibilizada em vídeos: uma abordagem utilizando ontologia, processamento de linguagem natural e descritores multimodais. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 13., 2012, Rio de Janeiro. **Anais**...2012.Rio de Janeiro, 2012. p. 20.
SOUSA, E. E.; DUQUE, C. G. Uso de ontologia para recuperação da informação disponibilizada em vídeos por meio de indexação multimodal. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 12., 2011, Brasília. **Anais**...2011.Brasília, 2011. p. 2730-2737.

SMIT, J. W. O documento audiovisual ou a proximidade entre as 3 marias. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 26, n. 1/2, p. 81-85, jan./jun. 1993.

Organização do conhecimento e *design* da informação: um estudo de convergência

Natalia Nakano
Universidade Estadual Paulista
natinakano@gmail.com

Mariana Cantisani Padua
Universidade Estadual Paulista
mariana.cantisani@gmail.com

Maria José Vicentini Jorente
Universidade Estadual Paulista
mjjorente@yahoo.com.br

1 Introdução

Sistematizar conhecimento, diante do volume da informação atualmente encontrada, em especial nos ambientes digitais, evidencia os campos da Organização do Conhecimento (OC) e do *Design* da Informação (DI), demonstrando interdisciplinaridades entre eles.

O assunto, entretanto, é ainda pouco discutido na Ciência da Informação (CI). A OC, de acordo com Hjorland (2008), pode ser entendida em seus sentidos restrito e amplo. No sentido restrito abarca atividades tradicionais realizadas em bibliotecas, arquivos e instituições de memória, tais como classificar, indexar e catalogar e tem como

disciplinas centrais a Biblioteconomia, a Documentação e a Ciência da Computação. No sentido amplo, a OC relaciona-se com a divisão do trabalho mental, ou seja, a OC como estrutura de áreas para pesquisa, organizações de universidades e outras instituições de produção e disseminação de conhecimento. Embora a Biblioteconomia e a CI sejam disciplinas dedicadas à OC no seu sentido restrito, é impossível falar em OC desconsiderando o seu sentido amplo, pois, atualmente, outras disciplinas também se preocupam com o estudo do conhecimento e sua organização. Pode se afirmar, portanto, que a OC ultrapassa os limites dos campos do conhecimento e permeia disciplinas nunca pensadas anteriormente. Hjorland (2008, 2013) reconhece a intersecção da OC com outros campos como a Ciência da Computação, Linguística, Psicologia, Ciências Sociais, Arquivística, Diplomática, e até a Arquitetura da Informação, porém não menciona as potencialidades do *Design* da Informação (DI) e sua contribuição para a OC.

Embora o DI tenha sido consolidado como disciplina recentemente, sua aplicação prática tradicionalmente se preocupa não só com a organização e acesso da informação, mas inclui um elemento indispensável na concepção sócio-cognitiva da OC: o elemento humano. O DI compreende o indivíduo inserido em um contexto social e cultural, e portanto, extrapola as questões cognitivas, especialmente depois da explosão informacional disponibilizada nos ambientes informacionais digitais. De acordo com Freitas, Coutinho e Waetcher (2013, p. 2):

> A efemeridade de informações nas últimas décadas, decorrentes das constantes transições tecnológicas, de valores, de atitudes e de necessidades ocorridas na sociedade, tornam a maneira de "tratar a informação" uma preocupação imprescindível no momento do projeto de *design,* e dobra, quando discutimos isso no Design da Informação.

Neste contexto se delineiam os problemas e objetivos de nossa investigação ao apresentar e discutir um dos pontos de convergência entre o *D*I e a OC: como as duas áreas se desenvolveram até a contemporaneidade? de que forma o *D*I pode inter-relacionar-se aos sistemas de OC de forma a torná-los mais eficientes e atraentes?

Dentre as inúmeras e valiosas contribuições de Hjorland para o campo da OC, está a defesa da visão sócio-cognitiva. O desenvolvimento desta visão parte de lacunas percebidas por Hjorland nos estudos de usuário dos anos de 1970-1980. Naquele

momento prevalecia na CI o modelo clássico de Recuperação da Informação (ROBERTSON, 1977). Trata-se de um modelo linear e horizontal, onde a informação é constituída de um lado, como instrumento de transmissão de sinais (informação como coisa) e, por outro lado, entendido como uma mensagem que altera as estruturas do receptor (conceito cognitivo de informação).

Hjorland critica essa abordagem, afirmando que a OC é social, e não apenas cognitiva. O autor defende, em contrapartida, a Abordagem Analítica de Domínio (AD) como uma perspectiva mais adequada para a análise de domínio na OC. Formulada no início dos anos 1990 como uma alternativa à visão cognitiva dominante na Biblioteconomia e na CI, a

> Análise de Domínio é um ponto de vista sociológico-epistemológico. A indexação de um dado documento deve refletir as necessidades de um dado grupo de usuários ou um dado propósito ideal. Em outras palavras, qualquer descrição ou representação de um dado documento é mais ou menos adequado para realização de determinadas tarefas. Uma descrição nunca é objetiva ou neutra, e o objetivo não é padronizar descrições ou fazer uma descrição uma vez para todos os grupos alvos diferentes. (Hjorland, 2008, p. 95)

De acordo com o autor, a AD é a única abordagem para a OC que examinou questões epistemológicas no campo, comparando as presunções feitas em diferentes abordagens para a OC e examinando as questões relativas à subjetividade e objetividade na OC. Importantes são as visões coletivas, compartilhadas por muitos.

Para Hjorland (2008), diferentes pontos de vista precisam de diferentes sistemas de organização. Da mesma maneira, qualquer sistema de OC é sempre enviesado em direção a uma posição filosófica que reflete aquele que o criou ou o encomendou. Não há plataforma neutra a partir da qual o conhecimento possa ser organizado. O profissional da informação necessita desenvolver uma abordagem conceitual e prática que esteja coerente com os objetivos e valores da organização para qual o sistema é desenvolvido. A abordagem sócio-cognitiva, além das questões cognitivas, inclui o contexto social em que o agente está inserido.

A maior crítica do autor relacionada ao campo da OC é que a partir de 1950, vimos a OC migrar de um sistema de informação (SI) para outro sem desenvolver uma base teórica e epistemológica norteadora do desenvolvimento desses sistemas e plataformas.

Pode-se afirmar que o *D*I, assim como as abordagens de análise de domínio para a OC, evoluiu paralelamente e de forma similar - ao incluir a visão social, o círculo hermenêutico e a iteratividade - que os sistemas de informação devem considerar. O *D*I pode ser entendido como a ciência que busca preparar informações para a utilização eficaz e eficiente em ambientes físicos e digitais, com os objetivos de:

> 1. Desenvolver documentos que sejam compreensíveis, precisos e rapidamente recuperáveis, além da sua transformação fácil em ações efetivas;
> 2. Projetar interações por meio de equipamentos que sejam naturais, fáceis e agradáveis. Isto envolve resolver os problemas do *design* da interface humano-computador.
> 3. Permitir que as pessoas consigam se orientar em um espaço tridimensional com facilidade e conforto, sendo esse espaço principalmente o espaço urbano, mas também o espaço virtual (HORN, 1999, p. 15).

Podemos presumir que diferentes abordagens e paradigmas co-existam em todos os domínios, e o *D*I surge como uma nova ciência e uma disciplina que traz consigo metodologias para o desenvolvimento de projetos visando organizar o conhecimento, estruturar a informação, e guiar a realização da busca e recuperação da informação em um sistema de informação (SI) com a finalidade de construção do conhecimento.

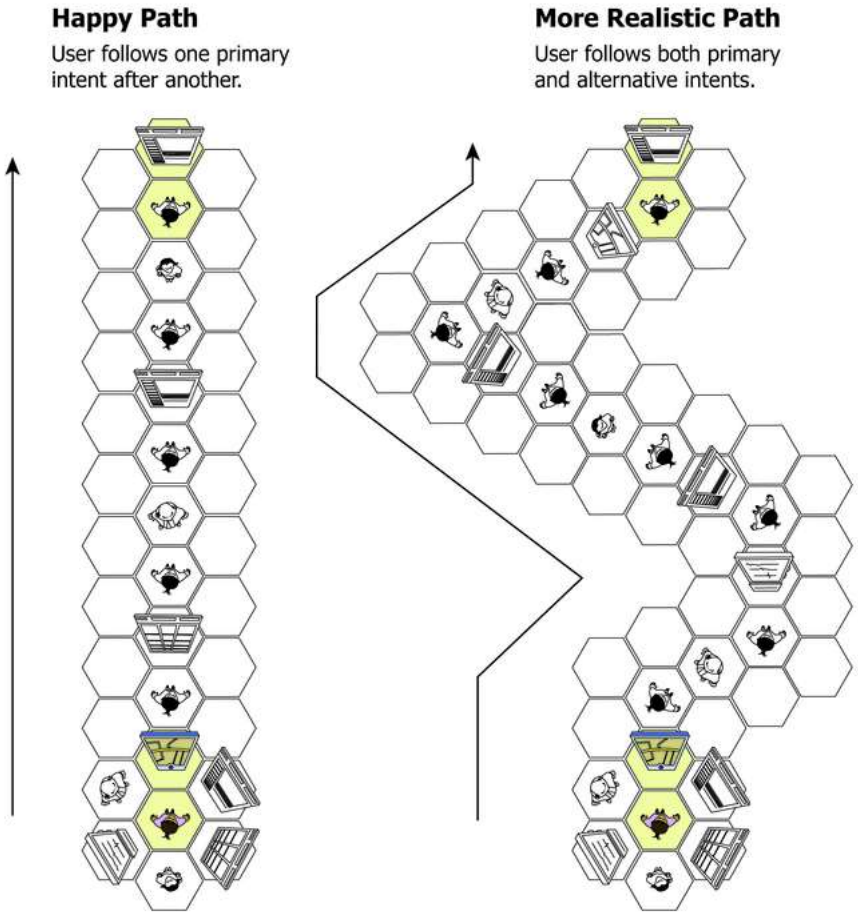
2 Modelos de interação com a Informação por meio do *Design* da Informação

Ao se considerar a OC, destaca-se aqui - como ponto de partida - a questão do círculo hermenêutico e suas funções iterativas de compreensão e interpretação. A base do círculo hermenêutico é o movimento, em que ocorre a compreensão do todo para as partes e das partes para o todo. Por essa razão, realiza-se no presente estudo uma correlação com os modelos de busca da informação e construção de conhecimento do *D*I.

De acordo com Palmer (1986), na hermenêutica fenomenológica é impossível separar o objeto, o contexto e o agente; Gadamer (1997, p. 516) aponta que é necessário "deixar as coisas se manifestarem como elas são, sem forçá-las com as nossas próprias categorias. Isso implica em uma reversão da direção usual. Não somos nós que apontamos para o que as coisas são; ao contrário, as coisas se mostram para nós".

Projetar/Desenhar experiências de interação e uso de informação em ambientes informacionais que sejam eficazes, requer não apenas uma consciência das características cognitivas do indivíduo, mas uma compreensão clara de como esses agentes percorrerão os caminhos na realização dessa interação com as informações, artefato e ambiente. Conforme mostrado na Figura 1 (WILLIS, 2011) o caminho percorrido pelo agente que interage com o SI não se dá de forma direta do sujeito para seu objetivo. Dessa maneira, o *designer* tem que considerar que o agente, com as suas diferenças sociais e culturais inserido em um determinado contexto, tomará diferentes caminhos. Assim, o sistema eficiente deve apresentar diferentes atalhos para o acesso à informação, e deve ter características resilientes, adaptando-se às diferentes necessidades do agente.

Figura 1 – Caminho projetado vs. caminho percorrido


Happy Path
User follows one primary intent after another.
More Realistic Path
User follows both primary and alternative intents.

Fonte: (WILLIS, 2011)

Já na década de 1970, Selle (1973) defendia que o *design* de ambientes artificiais deveria possibilitar e facilitar a compreensão do significado dos objetos e dos SI pelas pessoas. O *designer,* como agente configurador, deveria então considerar a realidade dinâmica da interação dos sujeitos com o sistema, desenvolvendo esse processo com o foco de criar uma experiência satisfatória, agradável e de fácil entendimento.

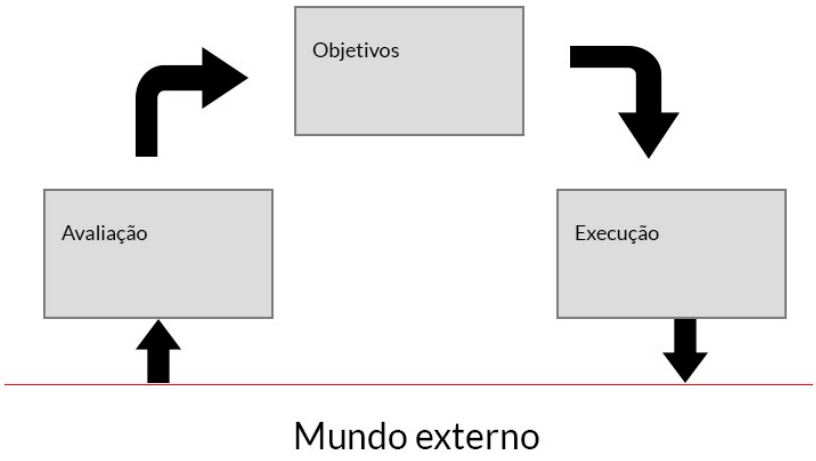
A concepção dos processos de interação com os ambientes de informação no *D*I tem evoluído de modelos simples e estáticos para modelos complexos e dinâmicos (RUSSEL-ROSE; TATE, 2013). Três modelos que moldaram o entendimento no processo de acesso à informação utilizado pelo *D*I são apresentados nesse estudo.

Norman (1988), provindo da computação, foi desenvolvedor da pesquisa pioneira conhecida como engenharia cognitiva. O modelo proposto pelo autor preocupa-se com

as maneiras como as pessoas interagem com a máquina; com os processos envolvidos no momento de interação dos atores com os sistemas; com as necessidades de interação dos indivíduos com os artefatos digitais para realização de suas tarefas.

No modelo cognitivo, baseado na performance de tarefa, o agente cria um modelo mental de metas, realiza uma sequência de ações, executa as ações, tem uma percepção da tarefa, interpreta e avalia os resultados (Figura 2).

Figura 2 – Modelo cognitivo proposto por Norman (1988)

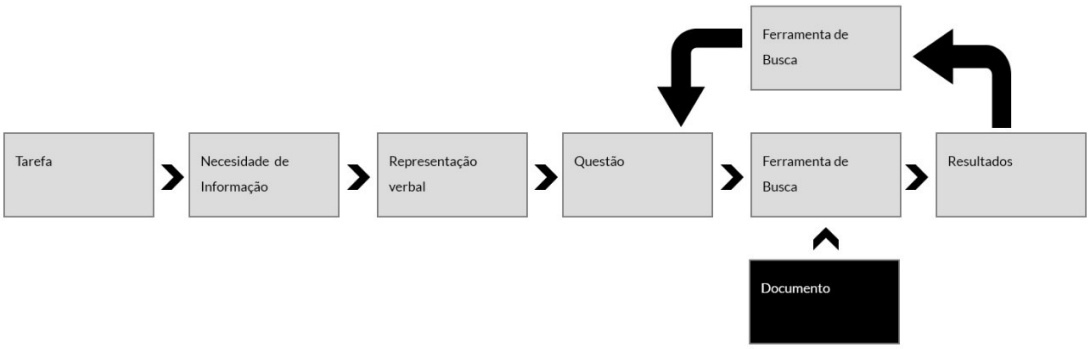


Fonte: (RUSSEL-ROSE; TATE, 2013, p.25, tradução nossa).

Fundamental neste modelo de Norman (1988) é que ele reconhece a importância do conhecimento do domínio: quanto maior o conhecimento do agente sobre o domínio, maior a eficácia das consultas, a precisão e a relevância dos resultados (RUSSEL-ROSE; TATE, 2013).

O modelo (Figura 3) desenvolvido por Marchionini (1995), conhecido como modelo padrão, veio a contrastar o modelo clássico da CI proposto por Robertson (1977). Marchionini (1995) defende que a busca da informação é uma forma de solução de problema, envolvendo um ciclo de quatro atividades: identificação; articulação sobre a necessidade da informação; formulação da questão; e avaliação dos resultados (STUCLIFFE; ENNIS, 1998).

Figura 3 – Modelo padrão proposto por Marchionini (1995)



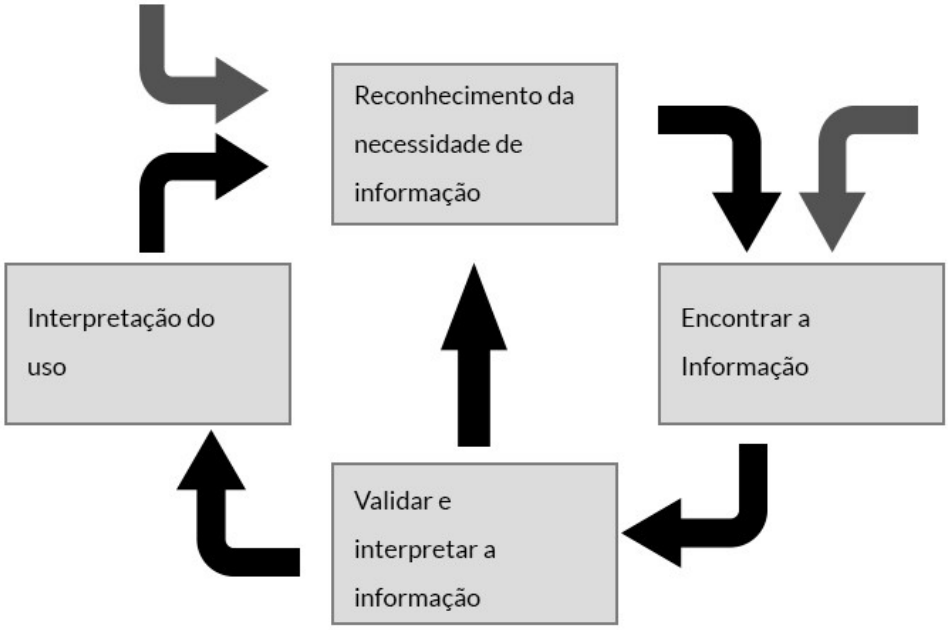
Fonte: (RUSSEL-ROSE; TATE, 2013, p.25, tradução nossa).

Diante da tarefa, a pessoa articula a sua necessidade informacional por meio de uma representação verbal, que por sua vez resulta em uma questão. Sua questão é então inserida em uma ferramenta de busca de documentos de uma coleção. A ferramenta apresenta um resultado, que se não for ideal, faz com que o indivíduo retorne à uma nova questão.

O modelo padrão é limitado considerando-se as relações dinâmicas relacionadas aos fatores humanos de interação com os sistemas de informação. Não considera que a necessidade de informação possa se transformar de acordo com os resultados recuperados. Além disso, o modelo também não considera os indivíduos em sua complexidade: os indivíduos são movidos cotidianamente por sensações, impressões e sentimentos, que influenciam as formas de interação com a informação, com o ambiente informacional e os SI.

A seguir, o modelo proposto por Blandford e Attfield (2010), conhecido como modelo da jornada da informação (Figura 4), derivou de estudos empíricos de comportamento informacional, tendo como atividades propostas o seguinte *framework*: reconhecimento da necessidade de informação; aquisição da informação; interpretação e validação da informação; e interpretação do uso.

Figura 4 – Modelo da jornada da informação proposta por Bland ford e Attfield (2010)


Reconhecimento da necessidade de informação
Encontrar a Informação
Validar e interpretar a informação
Interpretação do uso

Fonte: (RUSSEL-ROSE; TATE, 2013, p.25, tradução nossa).

Nele, as necessidades de informação dos agentes evoluem conforme a interação com a informação e assim novas metas são formuladas à medida que se adquire conhecimento de domínio. A busca por informação é, portanto, um processo iterativo e interativo no qual a resposta pode mudar a pergunta. Vale notar que esse modelo leva em conta o fenômeno *serendipity*, representado pelas flechas cinzas da Figura 4.

Percebe-se que esse modelo enfatiza a validação, interpretação e uso da informação - atividades chave, modeladoras da evolução da necessidade de informação, devido aos múltiplos pontos de entrada do modelo.

Contemporaneamente, o *D*I, além de considerar a visão sócio-cognitiva compartilhada com a OC, tem evoluído para a proposta de incluir a experiência do usuário (UX), o entendimento das emoções que influenciam processos como a atenção, reconhecimento e memória gerados pela interação dos agentes com os SI (DAMASIO, 2011;NORMAN, 2008).

A UX investiga questões de complexidade como a subjetividade, incerteza, transcendência e multidimensionalidade que extrapolam o sistema e sua relação com o agente, investiga o contexto, a cultura e as questões subjetivas de emoções dos usuários.

Assim como Hjorland (2008), não apenas considerando um indivíduo mas um grupo de indivíduos.

3 Considerações

Concluiu-se que, para uma sistematização otimizada do conhecimento, em especial nos ambientes digitais, a OC e o DI, são áreas interdisciplinares de bem vinda convergência nos planos conceituais, disciplinares e de práticas.

Tanto o *D*I quanto a OC consideram que um sistema eficiente se estrutura sobre questões complexas que incluem não apenas contexto e cultura dos sujeitos, mas também incertezas. No contexto da incerteza, um sistema eficiente deve ser resiliente para a realização satisfatória das tarefas. Além disso, o SI deve operar de forma intuitiva, ou seja, deve apresentar *affordance* relevante. Conceitos, métodos e teorias amplamente discutidas na DI, e que podem efetivar a OC.

Assim, há que se considerar a complexidade das inter-relações de variadas disciplinas com a OC e as contribuições teóricas e práticas que estas podem oferecer, como a interdisciplinaridade com o *D*I. Para a sustentação cientifica acadêmica de tal afirmação apresentaram-se neste artigo modelos conceituais do DI que demonstram consonância com o conhecimento produzido na OC.

Referências

BLANDFORD, A.; ATTFIELD, S. **Interacting with Information**. Morgan & Claypool, 2010. p. 29-39

DAMÁSIO, A. **E o cérebro criou o homem**. São Paulo: Companhia das Letras, 2011.

FREITAS, R.F.; COUTINHO, S. G.; WAETCHER, H. N. Análise de Metodologias em Design: a informação tratada por diferentes olhares In: **Estudos em Design | Revista (online).** Rio de Janeiro: v. 21 | n. 1, 2013. p.1-15.

HJORLAND, B**.** User-based and Cognitive Approaches to Knowledge Organization. **Knowledge Organization**, 40, 2013. p. 11-27.

HJORLAND, B. What is Knowledge Organization (KO)? **Knowledge Organization**. 35. n. 2/n. 3, 2008.

HORN, R. Information Design: Emergence of a New Profession. In: R. Jacobson (Org.). **Information Design**. Cambridge, Massachusetts: The MIT Press, 1999. p. 15-33.

GADAMER, Hans-Georg. A linguagem como *medium* da experiência hermenêutica. In: **Verdade e método I**. Trad. Flávio Paulo Meurer. Petrópolis: Vozes, 1997. p. 497- 631.
MARCHIONINI, G. **Information Seeking in Electronic Environments**. Cambridge: Cambridge University Press, 1995. p 27-60.

NORMAN, D. A. **The Psychology of Everyday Things***,* New York: Basic Books, 1988
NORMAN, D. A. **Design emocional: porque adoramos ou detestamos os objetos do dia-a-dia**. Rio de Janeiro: Rocco, 2008.

PALMER, R. **Hermenêutica**. Trad. Maria Luísa Ribeiro Ferreira. Lisboa: Edições 70, 1986.

ROBERTSON, S. E. Theories and models in information retrieval. **Journal of Documentations**, 33(2), 1977. p 126-148.

RUSSEL-ROSE, T.; TATE, T**. Designing the search experience**: the information architecture of discovery. Morgan Kaufmann, 2013. 302 p.

SELLE, G. **Ideologia y utopia del diseño. Contribución a la teoria del diseño industrial.** Barcelona: Editorial Gustavo Gill, 1973.

SUTCLIFFE, A. G.; ENNIS, M. Towards a cognitive theory of information retrieval. **Interacting with Computers**, 10, 1998. p 321-351.

WILLIS, D. **Intent Paths**. 2011. Disponível em: < https://www.flickr.com/photos/uxcrank/sets/72157626618631035/with/5730905682/> Acesso em: 5 mai. 2015.

Análisis de dominio de las dos primeras ediciones del capítulo ISKO-Brasil

Exequiel Fontans
Universidad de la República
exequiel151974@gmail.com

Mario Barité
Universidad de la República
mabarite@gmail.com

1 Introducción

Desde su fundación en el año 2007, el capítulo ISKO-Brasil se ha constituido en el vehículo de fomento e intercambio científico del área Organización del Conocimiento para Brasil. En las dos primeras ediciones de estos eventos según surge de las actas se presentaron más de 80 ponencias, que llevaron 123 firmas, provenientes de 38 instituciones de 9 países. En este marco es relevante analizar el dominio de la producción registrada en las actas del capítulo ISKO-Brasil (Brasilia, 2011 y Río de Janeiro, 2013), porque ello permitirá identificar las condiciones en las cuales el conocimiento científico se construye y socializa (Guimarães, Santos, Candido, & Pinho, 2014).

Smiraglia (2014, pág. 85) señala que “existe cierto nivel de confusión en la definición de un dominio, por lo que muchos autores han hablado de comunidades discursivas, disciplinas, colegios invisibles, e incluso ecologías, sin embargo todos ellos son tipos de dominios.” Y concluye: “Un dominio puede definirse sencillamente como una comunidad de académicos que trabajan en problemas de investigación que son percibidos de alguna manera como similares” (Smiraglia, 2014, pág. 85). Esto trae implícito que comparten una ontología, un conjunto de hipótesis comunes, una metodología de trabajo que aplican a los mismos problemas de investigación, y por tanto, un lenguaje común que les permite comunicarse (Smiraglia & Lee, 2012, pág. 114).

Por su parte el análisis de dominio, postulado para la Ciencia de la Información por Hjørland y Albrechtsen (1995), constituye un abordaje metodológico central en Organización del Conocimiento, que puede subdividirse en once enfoques metodológicos (Hjørland, 2002). Entre ellos se encuentran los estudios bibliométricos, en especial menciona el análisis de citas y de cocitación (Oliveira & Grácio, 2013). Para este trabajo, además del análisis de citas, se emplean el análisis de coocurrencia de palabras, y de colaboración. Estos abordajes bibliométricos son potenciados combinándolos con el análisis de redes sociales. Entre los numerosos antecedentes que aplican el enfoque bibliométrico y análisis de redes sociales, pueden mencionarse la caracterización de la investigación en Ciencia de la Información en Brasil reportada por Liberatore y Herrero-Solana (2013), el análisis del GT7del XIII ENANCIB 2012 (Guimarães, Moreiro González, & Alencar, 2012) y más recientemente el trabajo de Guimarães y colaboradores (2014), en donde analizan las actas de los congresos de Brasil, España y América del Norte para el periodo 2011- 2013.

En este marco, la presente propuesta tiene por objetivo contribuir al análisis de dominio del campo temático Organización del Conocimiento en Brasil mediante el uso combinado de herramientas bibliométricas y de análisis de redes sociales al corpus documental formado por las actas de los congresos organizados por el capítulo brasileño de ISKO. En particular se pretende a) identificar las temáticas abordadas, b) representar la colaboración personal, institucional y entre países, y c) analizar y visualizar las citas utilizadas.

2 Metodología

Para realizar este trabajo se realizaron los siguientes procedimientos metodológicos:

1 - Obtención del corpus documental

El corpus documental estuvo conformado por las dos actas de los congresos del capítulo ISKO-Brasil que se desarrollaron en Brasilia, 2011 y Río de Janeiro, 2013 (Guimarães & Dobedei, 2012, 2013).

Se volcaron los datos de las ponencias en una planilla de cálculo. Los datos recabados fueron: título de la ponencia, palabras clave, autores, filiación, año, citas bibliográficas.

Dado que las ponencias no cuentan con palabras clave se obtuvieron estas a través de un método indirecto: se identificaron y seleccionaron las palabras significativas de los títulos de las ponencias en el entendido de que representaban el contenido de las mismas.

2 – Depuración del corpus.

En esta etapa se normalizó el nombre de los autores, se tradujeron al portugués los títulos que estaban en otros idiomas (9 en inglés y 17 en español), y se normalizó el nombre de los autores citados.

3 – Análisis de los datos

Se confeccionaron las matrices de palabras clave, de autores, y autores citados, además de las matrices de atributos necesarias para el procesamiento posterior mediante el software UCINET 6 y NetDraw 2 (Borgatti, Everett, & Freeman, 2002).

Los análisis de métricas de análisis de redes sociales se hicieron a través de las medidas de centralidad. Este concepto fue propuesto por Freeman (1979), siendo los aportes de Newman (2000, 2001) muy relevantes para su aplicación a redes de colaboración. A continuación se describen someramente las principales medidas de centralidad.

Centralidad de Grado *(Degree):* Esta medida es la más común e intuitiva de las medidas de centralidad: parece razonable intuir que los nodos con mayor cantidad de relaciones con otros (los mejor conectados) tienden a ser más influyentes en la red.

Centralidad de Intermediación (*Betweeness*): si un nodo está en el camino más corto (la distancia geodésica, en teoría de grafos) entre otro par de nodos es seguro que tiene influencia entre la información que circula entre ese par de nodos.

Centralidad de Cercanía (*Closeness*): esta medida en una red proporciona un indicador de cuan cerca se encuentra un nodo en promedio del resto de nodos de la red. Los nodos con un alto índice de cercanía acceden muy rápido a la información que circula por la red. Si la red es de colaboración, los autores con un alto índice de cercanía pueden ser vistos como autores que acceden muy rápido a la información que circula por la red

y a su vez la información que ellos generan llega muy rápido al resto de autores de la red (Newman, 2000, 2004, 2010; Ye, Li, & Law, 2011). Si es una red de palabras clave, la cercanía puede interpretarse como afinidad temática.

Para el análisis de citaciones se parte de la presunción de que dos o más documentos que son citados juntos contienen una similaridad temática, al menos para el citante. En un corpus constituido por las actas de un congreso disciplinario como el abordado en este trabajo se espera que esta temática coincida con los tópicos del congreso. Por tanto se asume que los documentos altamente citados representan un núcleo temático que podría ser usado para identificar y visualizar las relaciones entre los autores citantes (Miguel, Moya-Anegón, & Herrero-Solana, 2006).

4 – Productos del procesamiento

a. Datos generales.
b. Análisis de palabras clave.
c. Análisis de colaboración.
d. Análisis de citas

3 Resultados

a – Datos generales

En la Tabla 1 se presentan los datos básicos de la red de colaboración. A continuación se destacan algunos datos relevantes. De los 83 documentos que componen el corpus, el 43% (n=36) son firmados por un solo autor. De los documentos realizados en colaboración el 64% (n=30) son firmados por dos autores. Por otra parte, solo el 18% (n=22) de autores no tiene trabajos en colaboración. Estos datos explican que la media de autores por documentos sea levemente mayor a dos (2,15).

Tabla 1. Datos básicos de la red de colaboración.

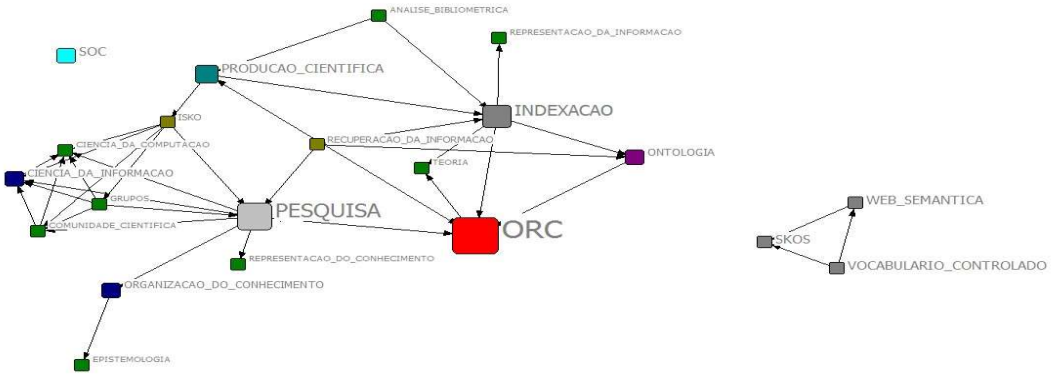
Número total de documentos	83	
Documentos realizados por un solo autor	36	43%
Documentos realizados en co-autoría	47	57%
Documentos realizados por dos autores	30	64%
Documentos realizados por tres autores	10	21%
Documentos realizados por más de tres autores	7	15%
Número total de autores	123	
Autores sin niguna colaboración	22	18%
Autores con alguna colaboración	101	82%
Autores con un único trabajo en colaboración	84	68%
Autores con dos trabajos en colaboración	17	14%
Autores con tres o más trabajos en colaboración	0	0%
Número total de colaboraciones	101	
Media de documentos por autor	0,67	
Media de autores por documentos en colaboración	2,15	
Número total de autores con tres o más trabajos en colaboración	0	0%

Fuente datos de la investigación.

b - Análisis de palabras claves

Del análisis de los títulos de las ponencias se obtuvieron 130 palabras clave. Para facilitar la visualización, se seleccionaron las palabras clave de grado mayor a 5 (Figura 2). En la Tabla 2 se presentan las medidas de centralidad de las 21 palabras que cumplen con este requisito. Las palabras están ordenadas por la medida de intermediación (*Betweeness*); puede decirse que las siete palabras clave que aparecen primero en este ranking constituyen el núcleo duro de la temática de los congresos del capítulo brasileño de ISKO (PESQUISA; ORC; INDEXACAO; ORGANIZACAO DO CONHECIMENTO; PRODUCAO CIENTIFICA; ISKO; RECUPERACAO DA INFORMACAO). Por su parte, la cercanía (*Closeness*) indica las palabras que están mejor conectadas con el resto de las temáticas (SKOS; WEB SEMANTICA; VOCABULARIO CONTROLADO; SOC; EPISTEMOLOGIA; REPRESENTACAO DA INFORMACAO; ANALISE BIBLIOMETRICA; REPRESENTACAO DO CONHECIMENTO; TEORIA; ONTOLOGIA).

Figura 2. Red de palabras claves de grado mayor a 5.



Fuente: datos de la investigación.

Tabla 2. Medidas de centralidad de red de palabras claves de grado mayor a 5.

Nombre	Rank	Grado	Grado Normalizado	Rank	Cercanía	Cercanía Normalizada	Rank	Interme-diación	Intermediación Normalizada
PESQUISA	1	9	0,155	13	703	0,195	1	65	0,117
ORC	4	5	0,217	12	706	0,199	2	25	0,154
INDEXACAO	2	7	0,116	9	711	0,192	2	25	0,083
ORGANIZACAO_DO_CONHECIMENTO	7	2	0,062	7	716	0,181	3	15	0,052
PRODUCAO_CIENTIFICA	5	4	0,085	10	710	0,188	4	14	0,023
ISKO	3	6	0,047	11	708	0,184	4	14	0,008
RECUPERACAO_DA_INFORMACAO	6	3	0,047	11	708	0,188	5	12	0,021
ONTOLOGIA	6	3	0,070	7	716	0,187	6	0	0,036
CIENCIA_DA_INFORMACAO	4	5	0,062	8	712	0,182	6	0	0,021
CIENCIA_DA_COMPUTACAO	4	5	0,039	8	712	0,179	6	0	0,000
COMUNIDADE_CIENTIFICA	4	5	0,039	8	712	0,179	6	0	0,000
GRUPOS	4	5	0,039	8	712	0,179	6	0	0,000
SKOS	7	2	0,047	1	764	0,117	6	0	0,000
WEB_SEMANTICA	7	2	0,047	1	764	0,117	6	0	0,000
VOCABULARIO_CONTROLADO	7	2	0,039	1	764	0,117	6	0	0,001
ANALISE_BIBLIOMETRICA	7	2	0,039	4	720	0,177	6	0	0,017
TEORIA	7	2	0,039	6	717	0,186	6	0	0,018
EPISTEMOLOGIA	8	1	0,039	2	731	0,171	6	0	0,006
REPRESENTACAO_DA_INFORMACAO	8	1	0,039	3	726	0,176	6	0	0,025
REPRESENTACAO_DO_CONHECIMENTO	8	1	0,039	5	718	0,179	6	0	0,012
SOC	9	0	0,062	1	774	0,118	6	0	0,003

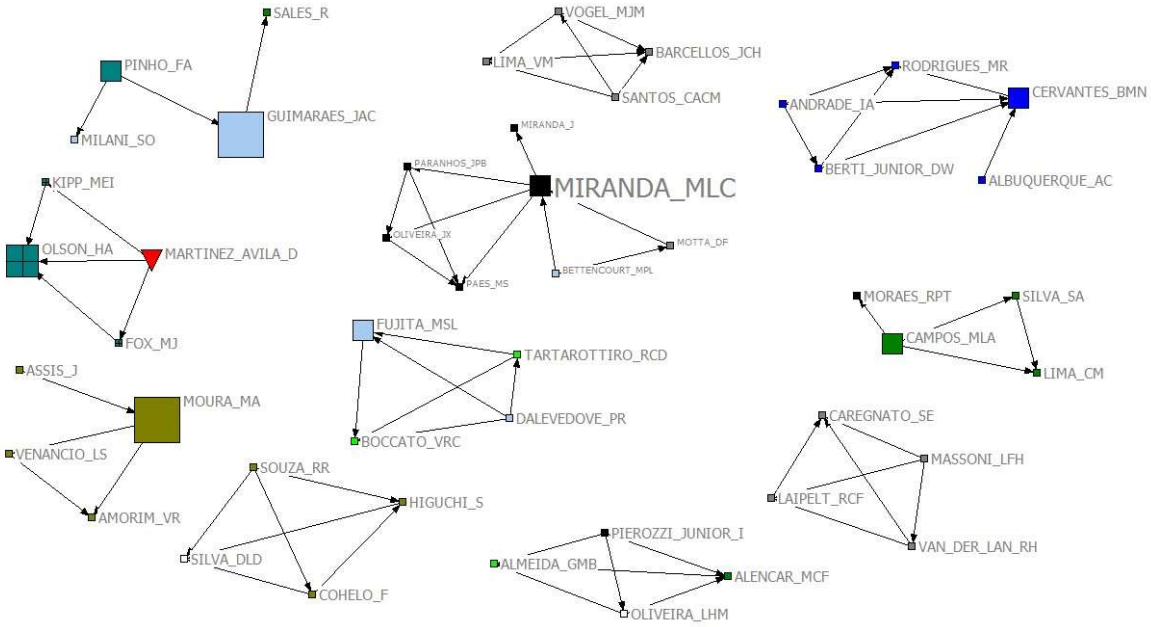
Fuente: datos de la investigación.

c – Análisis de colaboración

En la Figura 3 se presenta los grupos de colaboración formado por al menos 4 integrantes. Se pueden ver 9 grupos de 4 integrantes, uno de 5 (Cervantes BMN) y uno de 7 (Miranda MLC), aparecen aquí tres de los cuatro autores más productivos (Guimarães, Moura y Olson) que se listan en la Tabla 3. Además, puede verse la colaboración entre diferentes instituciones (color de los nodos) y países (forma de los nodos). De los 11 grupos detectados, 7 tienen algún tipo de colaboración interinstitucional. En la Figura 4 se presentan los autores sin colaboración y los grupos con 2 y 3 integrantes.

En la Tabla 4 se presenta la cantidad de autores por país; como es natural el 83% de los participantes son de Brasil, seguido por EEUU y España con 5% y 4% respectivamente. Los autores se distribuyen por instituciones como se observa en la Tabla 5; el 50% de los autores provienen de 7 instituciones brasileñas: UNIRIO; USP; UNESP; UFMG; UFF; UEL y UFRGS.

Figura 3. Red de colaboración formada por grupos de 4, 5 y 7 integrantes.



Fuente: datos de la investigación. Nota: El tamaño del nodo y de su etiqueta representa la cantidad de ponencias presentadas; la forma representa los países; el color la filiación.

Tabla 3. Detalle de los autores con más contribuciones.

Nombre	Cant. Art.	Colab.	País	Filiacion
GUIMARAES_JAC	4	2	BRASIL	UNESP
MOURA_MA	4	2	BRASIL	UFMG
OLSON_HA	3	2	EEUU	UWM
BARITE_M	3	2	URUGUAY	UDELAR

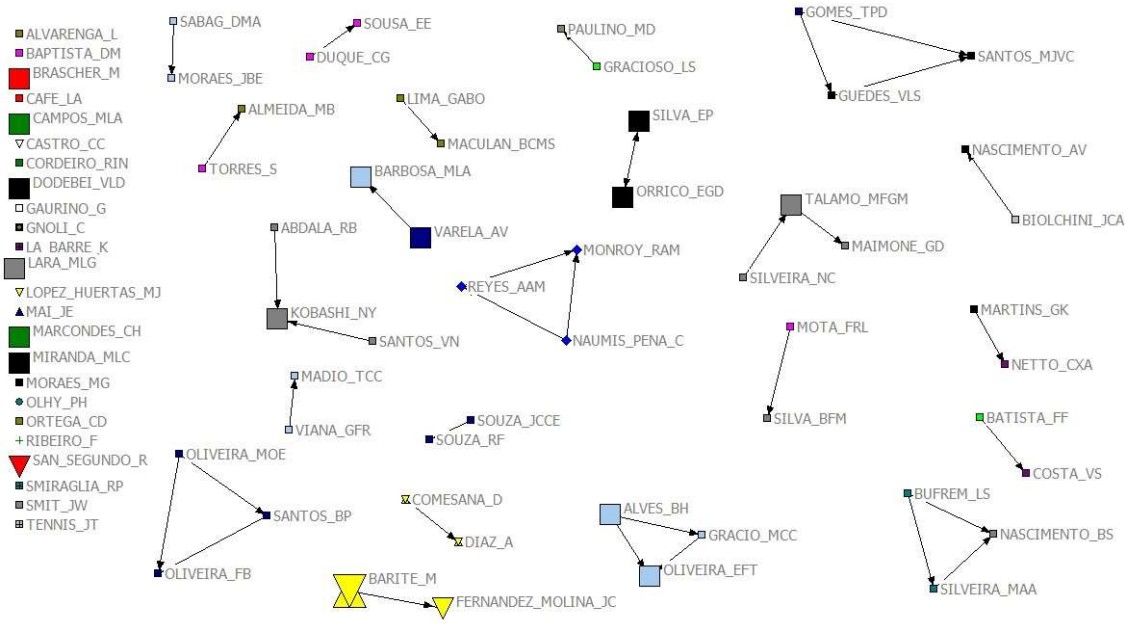
Fuente: datos de la investigación. Nota: Cant. Art.= Cantidad de artículos; Colab.= Colaboraciones.

Tabla 4. Detalle de los países participantes.

País	Cantidad	Porcentaje
ALEMANIA	1	1%
BRASIL	102	83%
DINAMARCA	1	1%
EEUU	6	5%
ESPANA	5	4%
ITALIA	1	1%
MEXICO	3	2%
PORTUGAL	1	1%
URUGUAY	3	2%
Total	**123**	

Fuente: datos de la investigación.

Figura 4. Red de colaboración formada por autores sin colaboración y grupos con 2 y 3 integrantes.



Fuente: datos de la investigación. Nota: El tamaño del nodo y de su etiqueta representa la cantidad de ponencias presentadas; la forma representa los países; el color la filiación.

Tabla 5. Detalle de las instituciones participantes.

Institución	Cant. Inst.	Cant. Aut.	Sub total	Porc.	Porc. Acum.
UNIRIO	1	13	13	11%	11%
USP	1	12	12	10%	20%
UNESP	1	11	11	9%	29%
UFMG	1	9	9	7%	37%
UFF	1	6	6	5%	41%
UEL; UFRGS	2	5	10	8%	50%
UFPA; UFSCAR; UNIVERSITY OF WISC(	3	4	12	10%	59%
FGV; IBICT; UDELAR; UFPE; UNAM; UN	6	3	18	15%	74%
EMBRAPA INFORMATICA AGROPECUARIA; UC3M; UFC; UFES; UFRJ; UFSC; UGR	7	2	14	11%	85%
ASSEMBLEIA LEGISLATIVA DO ESTADO DE MINAS GERAIS; CNEN; EMBRAPA MEIO AMBIENTE; IPEA; ISKO; THE UNIVERSITY OF WASHINGTON INFORMATION SCHOOL; UCB; UFAL; UFBA; UFPB; UNIVERSIDAD DE SALAMANCA; UNIVERSIDADE DO PORTO; UNIVERSITA DE PAVIA; UNIVERSITY OF COPENHAGEN; UNIVERSITY OF ILLINOIS AT URBANA CHAMPAIGN	15	1	15	12%	98%
Sin Dato	1	3	3	2%	100%

Fuente: datos de la investigación. Nota: Cant. Inst. = Cantidad de instituciones; Cant. Aut.=Cantidad de autores; Porc.= Porcentaje; Porc. Acum= Porcentaje acumulado.

d – Análisis de citas

Para el análisis de citas se contó la cantidad de citas para el primer autor de la obra citada (sin tomar en cuenta a los otros autores en caso de trabajos colectivos); igual criterio se estableció para los autores de las ponencias: se consideró “autor citante” al primer autor de la ponencia para aquellas contribuciones de autoría colectiva.

En la Tabla 6 se presenta el detalle de los autores más citados. Cinco autores, el 1% (Hjorland B; Dahlberg I; López-Huertas MJ; Fujita MSL y Smit JW), concentran el 9% de las citas (n=74). Si se consideran los autores que recibieron por lo menos 2 citas (22%, n=112), representan la mitad de las citas reportadas.

Como analizar este volumen de información (512 autores, 799 citas) es poco productivo como criterio de corte se tomaron aquellos autores que recibieron por lo menos 3 citas (n=53), que representan el 10% más citado del corpus, con el 35% del total de

citas. De la base de autores de ponencias con sus citas se tomaron sólo aquellos ponentes que habían citado a este núcleo de 53 autores más citados. Luego se realizó una matriz con los 85 autores obtenidos de esta operación. De estos autores, 23 son solo citantes (no recibieron citas), 37 son solo citados y el resto son citados y citantes. Se destaca que 60 ponencias citaron a alguno de estos autores de este núcleo duro. Ello implica que el 83% de los trabajos (sobre las 72 ponencias que tienen bibliografía), citó por lo menos uno de los 53 autores del núcleo más citado.

Tabla 6. Detalle de las citas.

Total general	Cant. Aut.	Acum	Porc. Acum.	Cant. Citas	Total citas	Citas Acum.	Porc. Acum.
HJORLAND B	1	1	0%	24	24	24	3%
DAHLBERG I	1	2	0%	23	23	47	6%
LÓPEZ-HUERTAS MJ	1	3	1%	10	10	57	7%
FUJITA MSL	1	4	1%	9	9	66	8%
SMIT JW	1	5	1%	8	8	74	9%
BARITÉ M; BEGHTOL C; CAMPOS MLA; GARCÍA GUTIÉRREZ AL; GARDIN J-C; GUIMARAES JAC	6	11	2%	7	42	116	15%
BOOK INDUSTRY STUDY GROUP; CAPURRO R; DODEBEI VLD; ISO; OLSON HA; SMIRAGLIA RP	6	17	3%	6	36	152	19%
FOSKETT AC; GUIZZARDI G; LANCASTER FW; MILLS J; RANGANATHAN SR	5	22	4%	5	25	177	22%
FOUCAULT M; GIL LEIVA I; GUARINO N; KOBASHI NY; LA BARRE K; MAI JE; MIRANDA MLC; OHLY HP; ORRICO EGD; SVENONIUS E; TENNIS JT	11	33	6%	4	44	221	28%
AITCHISON J; BUCKLAND MK; CASTELLS M; CHAUMIER J; CLARKE SGD; FARRADANE JEL; GARCIA MARCO FJ; GNOLI C; GONZALEZ DE GOMEZ MN; INGWERSEN P; ISKO; MARTÍNEZ-ÁVILA D; MOREIRO GONZÁLEZ JA; NOVAK JD; OLMEDA GÓMEZ C; PINHO FA; SHATFORD LS; SMITH B; SOERGEL D; TÁLAMO MFGM	20	53	10%	3	60	281	35%
Varios autores	59	112	22%	2	118	399	50%
Varios autores	400	512	100%	1	400	799	100%

Fuente: datos de la investigación. Nota: Cant. Aut.=Cantidad de autores; Acum.= Acumulado; Porc. Acum= Porcentaje acumulado; Cant. Citas= Cantidad de citas por autor; Citas Acum.= Citas acumuladas.

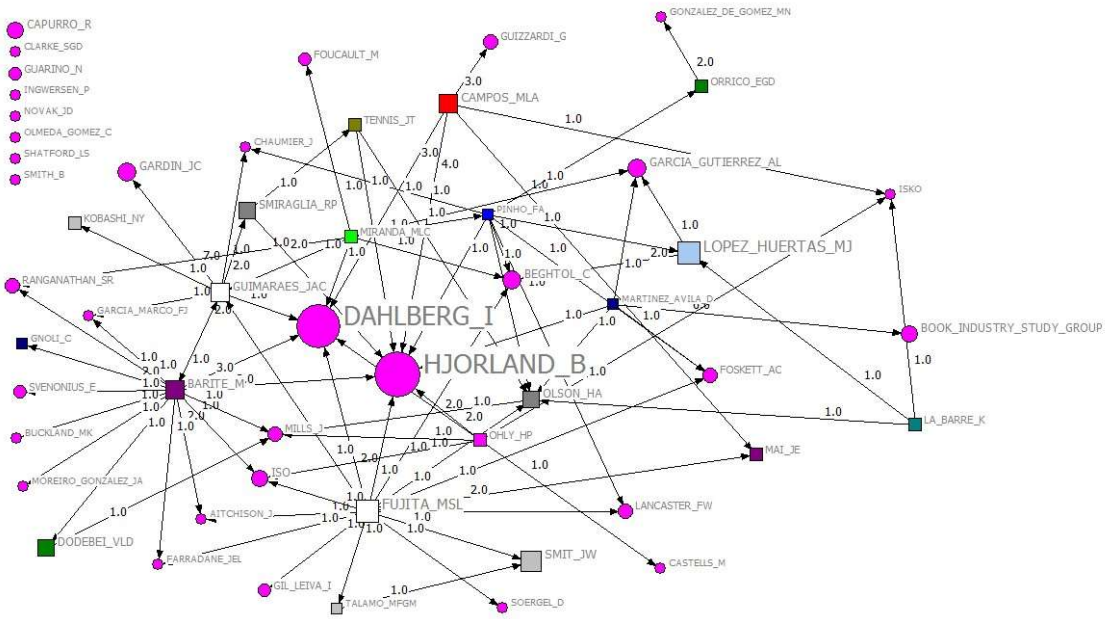
En la Figura 5 se presenta la red de citación del núcleo de 53 autores más citados del corpus. Un primer aspecto a considerar es que el componente gigante de la red incluye a todos los autores que participaron de los congresos del capítulo ISKO-Brasil. Ello indica que para cada nodo del componente gigante existe un camino que lo relaciona con cualquier otro nodo de la red. En otras palabras, en esta figura puede verse la

configuración de una comunidad de investigadores locales[2] que comparten una red de citación que incluye a los referentes internacionales de la disciplina.

En la Figura 6 se eliminaron los autores que no participaron de los congresos analizados. El resultado es una red de citación que incluye en su componente gigante a 19 de los 20 autores pertenecientes al núcleo de autores más citados. Esto refuerza la idea expresada antes de la existencia de una comunidad local que es citada por el resto de los participantes de los congresos, pero además se cita mutuamente.

En la Figura 7 se incluyeron a la red de la Figura 6 los autores que citan al núcleo de autores más citados. Por su parte, Los colores de la figuras 6 y 7 dan cuenta de una alta colaboración interinstitucional.

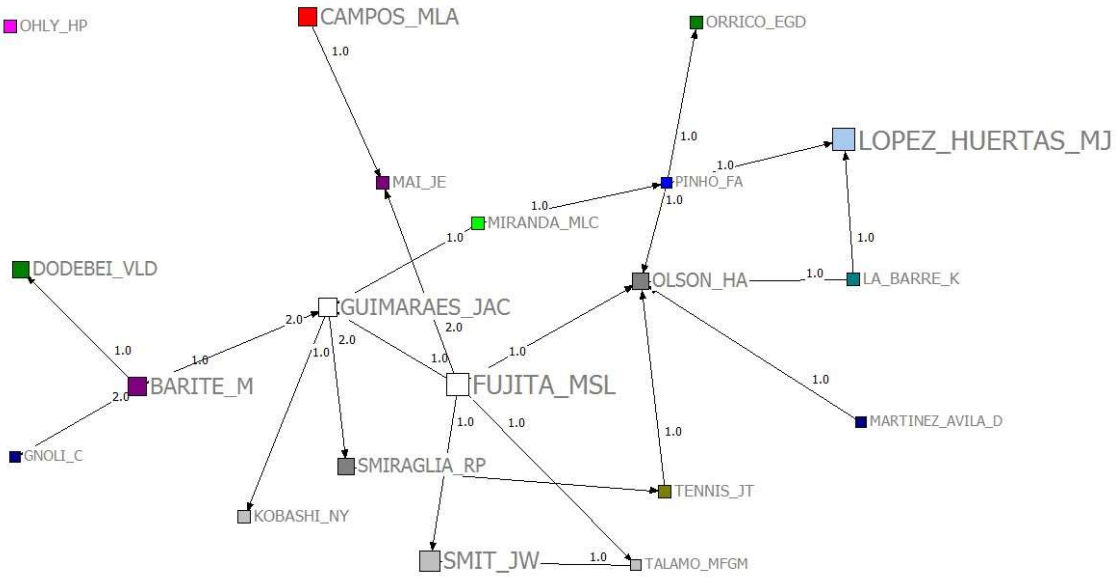
Figura 5. Red de citación de los 53 autores más citados del corpus.



Fuente: datos de la investigación. Nota: El tamaño del nodo y de su etiqueta representan la cantidad de citas recibidas, incluida las autocitas; la punta de la flecha señala al autor citado y el número, la cantidad de veces que lo citó; en círculo se representan los autores que no presentaron ponencia y en cuadrado los que sí lo hicieron; los colores de los cuadrados identifican instituciones.

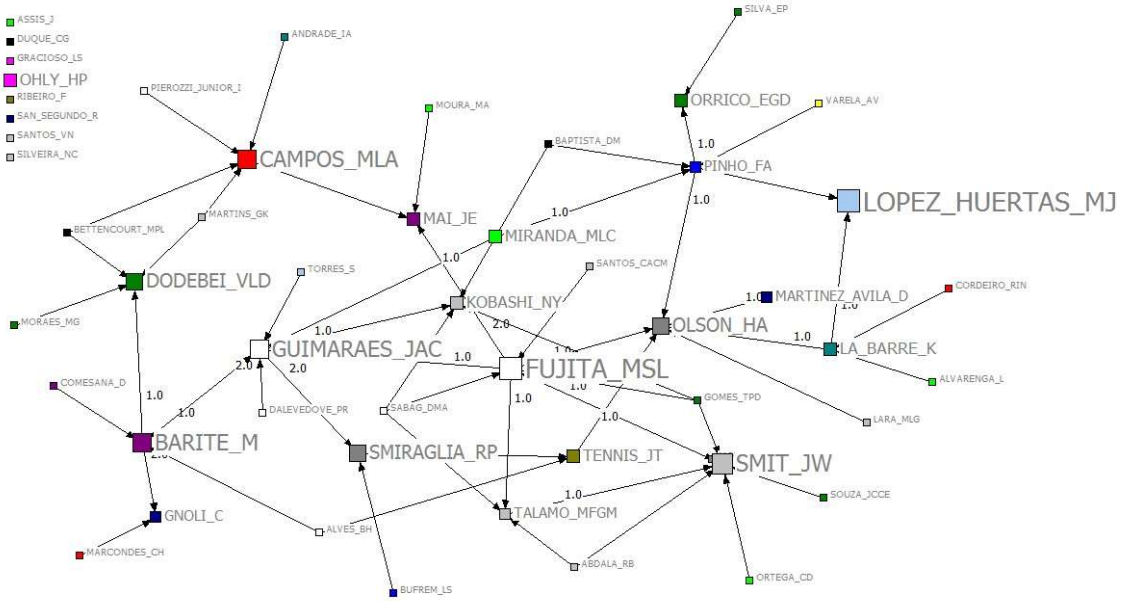
[2] En sentido laxo, dado que el 7% de ponentes no son de Brasil.

Figura 6. Red de citación de los autores que participaron de los congresos del capítulo ISKO-Brasil pertenecientes al núcleo de autores más citados.



Fuente: datos de la investigación. Ver nota de la figura 5.

Figura 7. Red de citación de los autores que participaron de los congresos del capítulo ISKO-Brasil.



Fuente: datos de la investigación. Ver nota de la figura 5.

4 Discusión y conclusiones

Este trabajo analizó las actas de los congresos del capítulo brasileño de ISKO en sus dos ediciones 2011 y 2013 mediante la utilización de técnicas bibliométricas y análisis de redes sociales. Los estudios bibliométricos constituyen un enfoque coherente y objetivo para analizar y caracterizar un dominio científico (Hjørland, 2002; Oliveira & Grácio, 2013); sin embargo sus resultados deben ser tomados con cautela si no se combinan con métodos cualitativos. En este marco, este trabajo es exploratorio y busca describir un fenómeno complejo con una herramienta válida pero con limitaciones intrínsecas.

Se identificó un núcleo conformado por siete palabras claves que representan las temáticas tratadas en estos dos eventos a saber: PESQUISA; ORC; INDEXACAO; ORGANIZACAO DO CONHECIMENTO; PRODUCAO CIENTIFICA; ISKO; RECUPERACAO DA INFORMACAO; y otro núcleo de palabras clave con alto índice de cercanía, lo que indica que están mejor conectadas con el resto de las palabras del estudio (SKOS; WEB SEMANTICA; VOCABULARIO CONTROLADO; SOC; EPISTEMOLOGIA; REPRESENTACAO DA INFORMACAO; ANALISE BIBLIOMETRICA; REPRESENTACAO DO CONHECIMENTO; TEORIA; ONTOLOGIA). El hecho de que las palabras clave se obtuvieran de una manera indirecta (de los títulos de las ponencias) es una limitación que debe ser tomada en cuenta.

Mediante el método de análisis de copublicación se identificaron las redes de colaboración que permitieron visualizar principalmente la colaboración interinstitucional e internacional. Lo breve del corpus documental no permitió identificar comunidades con este método. La coautoría, entendida como la firma conjunta de un trabajo, sigue siendo uno de los métodos más utilizados para estudiar la colaboración, sin embargo estos abordajes tienen sus limitaciones que han sido reportadas, entre otros, por Katz y Martin (1997).

El análisis de citas[3] permitió identificar y visualizar una comunidad de autores que conforman el núcleo de autores más citados en el corpus documental analizado. Este núcleo está integrado por referentes de la disciplina a nivel internacional y por autores locales, lo que podría estar dando cuenta del importante desarrollo de la OC en Brasil. Además, a nivel local se constata que existe colaboración interinstitucional, aunque no se profundizó en esta parte del análisis como para poder realizar una descripción con mayor precisión.

Referencias

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet 6 for Windows: Software for Social Network Analysis* (p. 302). Harvard: Analytic Technologies.

Freeman, L. C. (1979). Centrality in Social Networks Conceptual Clarification. *Social Networks*, *1*(3), 215-239.

Guimarães, J. A. C., & Dobedei, V. [Organizadores]. (2012). *Desafios e perspectivas científicas para a organização e representação do conhecimento na atualidade* (p. 285). Marília: ISKO-Brasil; FUNDEPE.

Guimarães, J. A. C., & Dobedei, V. [Organizadores]. (2013). *Complexidade e organização do conhecimento: desafios de nosso século* (p. 310). Río de Janeiro: ISKO-Brasil; FUNDEPE.

Guimarães, J. A. C., Moreiro González, J. A., & Alencar, M. F. (2012). A análise documental no universo científico dos ENANCIBS: elementos para uma análise de domínio. En *XIII Encontro Nacional de Pesquisa em Ciência da Informação - XIII ENANCIB 2012* (p. 17). Río de Janeiro.

Guimarães, J. A. C., Santos, A. Y. dos, Candido, G. G., & Pinho, F. A. (2014). A dimensão temática da pesquisa em organização do conhecimento: uma análise de domínio dos congressos nacionais e regionais da ISKO (Brasil, Espanha e América do Norte). *Scire*, *20*(2), 19-25.

Hjørland, B. (2002). Domain analysis in information science: Eleven approaches – traditional as well as innovative. *Journal of Documentation*, *58*(4), 422-462. doi:10.1108/00220410210431136

[3] Tomando en cuenta lo breve del corpus documental, para este trabajo no se eliminaron las autocitas, lo que seguramente influye sobre el resultado del análisis.

Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis. *Journal of the American Society for Information Science*, *46*(6), 400-425. doi:10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1-18. doi:10.1016/S0048-7333(96)00917-1

Liberatore, G., & Herrero-Solana, V. (2013). Caracterización temática de la investigación en Ciencia de la Información en Brasil en el período 2000-2009. *Transinformação*, *25*(3), 225-235. doi:10.1590/S0103-37862013000300005

Miguel, S., Moya-Anegón, F., & Herrero-Solana, V. (2006). El análisis de co-citas como método de investigación en Bibliotecología y Ciencia de la Información. *Investigación Bibliotecológica*, *21*(43), 139-155.

Newman, M. E. J. (2000). Who is the best connected scientist? A study of scientific coauthorship networks. En *Complex Networks* (Vol. 370, pp. 337-370). doi:10.1103/PhysRevE.64.016132

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 404-9. doi:10.1073/pnas.021544898

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(1), 5200-5205. doi:10.1073/pnas.0307545100

Newman, M. E. J. (2010). *Networks: An introduction* (p. 772). New York: Oxford University Press.

Oliveira, E. F. T., & Grácio, M. C. C. (2013). Studies of Author Cocitation Analysis: a approach for Domain Analysis. *IRIS: Informação, Memória e Tecnologia*, *2*(1), 12-23.

Smiraglia, R. P. (2014). *The Elements of Knowledge Organization (p. 101). New York: Springer.*

Smiraglia, R. P., & Lee, H. L. (Eds.). (2012). *Cultural Frames of Knowledge* (p. 158). Würzburg: Ergon-Verlag.

Ye, Q., Li, T., & Law, R. (2011). A coauthorship Network Analysis of Tourism and Hospitality Research collaboration. *Journal of Hospitality & Tourism Research*, *37*(1), 51-76. doi:10.1177/1096348011425500

Ensino em organização e representação do conhecimento: linguagens e normas

Cibele Araújo Marques dos Santos
Universidade de São Paulo
cibeleac@usp.br

Eduardo de Abreu de Jesus
Universidade de São Paulo
eduardo.jesus@usp.br

João Ricardo de Luca
Universidade de São Paulo
joao.luca@usp.br

1 Introdução

As ferramentas utilizadas para organização e representação da informação contida em documentos impressos ou digitais compõem os sistemas de organização do conhecimento, constituídos pelas linguagens documentárias como vocabulários controlados, listas de cabeçalhos de assuntos, sistemas de classificação e tesauros, bem como, os outros tipos de esquemas: as taxonomias e ontologias (CARLAN e BRÄSCHER, 2011). Tais sistemas, especialmente as linguagens documentárias de

acordo com Campos (2006), podem ser definidos pelo controle do vocabulário e pela dinâmica estabelecida de termos relacionados semântica e genericamente cobrindo um domínio específico do conhecimento.

Estas linguagens abrangem aportes teóricos e metodológicos que incluem a linguística, a lógica, a indexação, a terminologia e a construção de tesauros e vocabulários controlados, bem como o uso de sistemas de classificação e listas de cabeçalhos de assunto. A internet e a web semântica permitiram o surgimento de formatos e padrões como esquemas XML, o SKOS (*Simple Knowledge Organization System*) e as ontologias.

Assim, a interdisciplinaridade e complexidade destas ferramentas torna imperativo estudar estes recursos e repensar o ensino das linguagens documentárias e demais sistemas de organização do conhecimento para os profissionais da informação.

Em um projeto desenvolvido para identificar as práticas de ensino e estudar as tendências dos formatos, padrões e normas destes sistemas foi possível observar que a literatura da organização e representação do conhecimento apresenta uma lacuna na pesquisa sobre metodologias de ensino-aprendizagem das linguagens e esquemas de organização da informação.

2 Metodologia

Foi realizada a pesquisa bibliográfica nas bases de dados internacionais da área de Ciência da Informação LISA e ISTA. Foram pesquisadas também as bases multidisciplinares Web of Science e SCOPUS, mas nestas não foram encontrados artigos. Nas bases nacionais foram pesquisadas a BRAPCI, onde foram encontrados apenas 3 artigos sobre ensino de indexação e na SCIELO, que não trouxe resultados. A estratégia de busca utilizou palavras-chaves como “linguagens documentárias”, “classificações bibliográficas”, “tesauro” e “vocabulário controlado” cruzando com “ensino” em português e inglês, nas palavras do título, texto e resumo e na base LISA pesquisou-se utilizando o tesauro com estratégia semelhante. Alguns artigos foram selecionados, principalmente da base LISA para análise.

3 Características dos Sistemas de Organização do Conhecimento

O ensino de ferramentas de organização do conhecimento como as linguagens documentárias demanda contribuições interdisciplinares teóricas e práticas que apresentam certa complexidade, considerando as características dos sistemas e esquemas existentes.

Para Leise (2008), que apresenta artigo bem didático, um vocabulário controlado se diferencia de uma lista de indexação por diversos fatores: tempo, forma e finalidade de elaboração, e as características essenciais de um vocabulário controlado são as relações entre seus termos, como o controle de sinonímia, a estrutura hierárquica e a relação associativa. O autor adota a terminologia utilizada pela norma ANSI/NISO Z39.19 (Vocabulário Controlado) para se referir às linguagens documentárias e como a norma inclui as taxonomias neste escopo (NATIONAL INFORMATION STANDARDS ORGANIZATION, 2010). Também alerta para a diversidade terminológica quanto às linguagens e a dificuldade de definição e conceituação dos termos da área. O artigo define e explica utilizando quadros, além do vocabulário controlado, o anel de sinônimos, o controle de autoridade, a taxonomia e os tesauros. Traz também o conceito de faceta e sua importância para a categorização e estruturação de um vocabulário controlado.

Hedden (2008) também define de forma didática os conceitos de Vocabulário Controlado, como uma lista de termos para a elaboração da indexação e da categorização, sendo controlado porque define o termo a ser utilizado na indexação e sua importância é visível, mais facilmente, quando tratamos da recuperação de informações. Para a autora, o tesauro contempla além do conceito de vocabulário controlado, uma estrutura hierárquica que fornece informações sobre cada termo e seus relacionamentos. Considera a taxonomia como um vocabulário controlado com estrutura hierárquica, mas isenta de relações de equivalência e sinonímia e outros requisitos do tesauro tradicional.

A taxonomia de navegação para Leise (2008) corresponde a um vocabulário controlado para organizar websites ou outros conteúdos digitais, criando-se etiquetas, para categorias de navegação e diferentes níveis de hierarquização, que devem ser

designadas para refletir um contexto particular, como já acontece com os vocabulários, e devem ser usadas em conjunto com sistemas de gestão de conteúdo de websites.

Enquanto o design e a construção de vocabulários controlados têm sido apresentados em nível técnico detalhadamente ao longo das últimas décadas, o nível metodológico tem sido negligenciado, de forma que pesquisadores têm argumentado que a elaboração dessas linguagens precisa se basear nas análises e entendimentos dos contextos em que se inserem (MAI, 2008).

Informações e orientações sobre gestão dos sistemas de organização do conhecimento são importantes, pois a literatura sobre manutenção e revisão de linguagens documentárias e vocabulários controlados é escassa e dispersa, e também são poucos os trabalhos sobre o ensino desta atividade.

A manutenção de um vocabulário controlado, como enfatiza Leise (2008), deve ser constante, considerando a adição/supressão/ revisão de termos, reorganização das hierarquias, modificação, adição e deleção da etiqueta da faceta, atualização, controle de sinônimos, mudanças no processo de controle e definição de propriedade do vocabulário.

Além disso, Leise (2008) destaca que embora existam muitos conceitos e ideias novas a serem aprendidos na área, muitos indexadores consideram a elaboração de um vocabulário controlado como uma extensão de sua atividade de indexação.

A construção de vocabulários controlados está relacionada à indexação, segundo Hedden (2008), no que diz respeito a seus objetivos. Para a autora, as habilidades necessárias para a indexação são as mesmas habilidades necessárias para o desenvolvimento das linguagens e vocabulários, e incluem decidir sobre a melhor palavra para um conceito, criação de termo geral e específico em uma hierarquia, definição de entrada principal e os termos preferidos/não preferidos, referência e relação entre os termos.

A questão que se coloca aqui é: como desenvolver estas habilidades para que os estudantes de graduação e os profissionais graduados possam aprender as competências necessárias para criação e gestão de sistemas de organização do conhecimento. Também existe a necessidade de aprender a selecionar linguagens documentárias quando necessário, compatibilizar linguagens, desenvolver e ou utilizar

softwares e formatos, aprender as normas internacionais para vocabulários controlados e os esquemas para interoperabilidade.

4 Ensino e normas

A literatura pesquisada trouxe artigos sobre a tipologia das linguagens documentárias, suas definições e estrutura, usos, aplicações e normas, mas foram poucos os trabalhos sobre o ensino dessas linguagens.

Para Caro-Castro (2006), nas linguagens documentárias o que importa é o que está codificado lexicamente, não gramaticalmente, razão pela qual se eliminam as formas flexionadas dos termos substantivos escolhidos para descritores, assim estes são unidades das linguagens com função de meta-representação e que não mais apresentam ambiguidade, por terem sido normalizadas. Isto significa que foram controladas as relações entre conceito e descritor. Para isso, são utilizados métodos que estabelecem relações de equivalência e especifica-se o significado dos descritores mediante códigos e notas de alcance. A autora lembra também que a organização de um campo de conhecimento em um tesauro, por exemplo, favorece sua compreensão ao indicar não apenas conceitos representados, mas as inter-relações que os unem.

Esteban-Navarro (1995) reforça, em trabalho apresentado em evento sobre o ensino de linguagens documentárias, que bibliotecários devem ter conhecimentos de princípios para manutenção destas linguagens nas unidades de informação e desenvolvem estas competências em curso com apresentação da metodologia de aplicação dos princípios básicos de terminologia para gestão de tesauros, incluindo fundamentos científicos e exercícios práticos. Considera como aportes teóricos necessários: a terminologia, estudos lexicográficos e terminológicos em área específica, fontes de trabalho terminográfico de localização, identificação e seguimento de novos termos e conceitos mediante busca, seleção e recuperação de documentação técnico-científica representativa, atual e explícita, atribuição de códigos de classificação aos conceitos dos registros terminológicos, gestão de sistemas conceituais pelo uso e gestão de linguagens documentárias, terminologias normalizadas mediante implantação de sistemas de informação científicos, publicação e difusão de documentos secundários e controle da difusão de termos através de estudos bibliométricos.

A presença das linguagens documentárias nos estudos de graduação, segundo Espelt (1995), que apresentou trabalho no mesmo evento, não se limita à definição, características básicas e aplicação prática das linguagens mais utilizadas, mas também devem focar na terminologia e na análise de facetas. Para a autora, é importante que os profissionais estejam familiarizados com as teorias, princípios e diretrizes terminológicas. O aluno deve aprender a ser consciente que sua função se baseia na existência de vários níveis de linguagem e seu trabalho consiste em mover-se entre a terminologia dos documentos, a terminologia do sistema de informação e a terminologia dos usuários.

A normalização dos termos facilita o tratamento dos documentos, lembra Esteban-Navarro (1995), e favorece a progressiva homogeneidade do vocabulário, a recuperação e a difusão da informação, permitindo que autores, profissionais da informação e usuários utilizem os mesmos termos para designar os mesmos conceitos. Também a análise de como as definições dos conceitos e o controle dos sistemas terminológicos e conceituais podem ser aplicados à gestão de linguagens documentárias, já que fornece vocabulários baseados nos princípios de estruturação e relação conceitual, que comunicam a situação precisa de um conceito dentro de um campo temático. A linguagem documentária aparece como normalizadora nos níveis morfológico, sintático e semântico para garantia da estabilidade e eficiência na organização da informação.

O autor expõe a importância do ensino da definição de conceitos terminológicos e de suas regras de elaboração baseadas na norma ISO 704-1987. A aprendizagem das técnicas de construção de definições permite adquirir um recurso fundamental para manter e construir essas linguagens, já que o estabelecimento de relações entre os termos deve ser precedido por fixação restrita da extensão dos conceitos. Para obter univocidade são utilizadas diversas técnicas de desambiguação ou eliminação da sinonímia ou polissemia que de acordo com Espelt (1995), incluem as operações lógicas entre conceitos e as operações aplicadas na definição são instrumentos para determinar sinonímia e polissemia entre termos, para estabelecer especificidade da linguagem e aplicação correta.

Também no controle da relação entre os conceitos, o método de construção dos sistemas de conceitos é a melhor estratégia no entender de Esteban-Navarro (1995) para conhecer o sentido exato do conceito, já que estes se apresentam em relação aos outros

dentro do campo conceitual, e como fundamento básico das estruturas lógico-semânticas das linguagens documentárias que refletem a sistematização do conhecimento existente em um dado campo, segundo seu estado de desenvolvimento e segundo a tendência da taxonomia dominante. Apresenta a análise dos principais critérios de agrupamento estrutural ou de construção das taxonomias para mostrar como as relações genéricas e as ontológicas são as hierarquias, indicando também os diversos sistemas de representação gráfica que podem ser utilizados.

É importante o ensino da aplicação destes tipos de relações em casos concretos, para mostrar os três tipos de sistemas de conceitos documentários: os hierárquicos, os partitivos e os mistos, e critérios de agrupamento de conceitos que utilizam temas, disciplinas ou facetas, bem como as regras para apresentação dos termos nos tesauros de acordo com a normativa internacional (ESTEBAN-NAVARRO, 1995).

Para Espelt (1995), deve-se também considerar que o conhecimento das características dos textos científicos e técnicos faz parte do preparo para análise do conteúdo dos documentos. Além disso, considera conveniente o conhecimento dos sistemas de conceitos e os métodos aplicados na definição de termos para a correta aplicação das linguagens documentárias na indexação e recuperação da informação. A autora lembra que a equipe de elaboração de um tesauro deve ser composta por especialistas da área temática, terminólogos e documentalistas. Do mesmo modo, destaca a importância de desenvolver exercícios com os alunos como análise de textos científico-técnicos, definições de termos com vistas a se tornarem descritores, listas de termos com relações e elaboração de sistemas de conceitos. Igualmente, as classificações facetadas são importantes para a análise do conteúdo dos documentos e também para a qualidade da indexação, e reforça ainda que a literatura nesta área de ensino é pouca e os manuais de indexação são muito gerais sobre isso, de forma que o ensino de linguagens documentárias e de indexação devem ter pautas específicas.

Moreiro-Gonzalez (1995), salienta que, ao mesmo tempo, existe uma quantidade significativa de alunos com pouco preparo em gramática, ou até mesmo com dificuldades para a interpretação do texto que analisam. Por isso, propôs que a aprendizagem ocorra em fases, em um primeiro momento com a introdução de estudos linguísticos preparatórios, o ensino dos sistemas de classificação, das técnicas de indexação e

resumo tendo nesta fase um componente aplicativo predominante. Em uma segunda fase deve ocorrer o aprofundamento do tema, com a inclusão de uma variedade de elementos e conhecimento que permitam transformações representativas no processo.

Clarke e Zeng (2012) mostram que os princípios para a construção de um tesauro mudaram radicalmente da norma ISO 2788 para a ISO 25964, publicada em 2011, o que consideramos que se justifica pelas mudanças ocorridas no mundo com a tecnologia da informação, com a necessidade da interoperabilidade para a Web Semântica, onde a distinção entre termo e conceito é fundamental, contribuição importante implementada na nova norma. Assim, o SKOS (*Simple Knowledge Organization System*), recomendação do W3Consortium foi desenvolvido para a publicação de vocabulários e tesauros na Web.

Em comparação com outras soluções o SKOS oferece uma alternativa cuja aplicação é simples e rápida (PASTOR-SÁNCHEZ; MARTÍNEZ-MÉNDEZ; RODRÍGUEZ-MUÑOZ, 2012). Os autores informam que o SKOS se define formalmente como uma ontologia que permite representar qualquer tipo de sistema de organização do conhecimento mediante RDF. Seu âmbito de aplicação se estende praticamente a totalidade dos vocabulários controlados: classificações, tesauros, cabeçalhos de assunto, taxonomias, glossários, etc. No SKOS, os elementos de um vocabulário se representam mediante conceitos entre os quais se estabelecem relações semânticas, hierárquicas, associativas, de idioma, de sinonímia, de alternativas de acesso a um conceito através de etiquetas.

5 Considerações finais

Esta pesquisa permitiu identificar uma lacuna que existe na literatura, principalmente no Brasil, sobre metodologia e práticas de ensino de linguagens documentárias, a importância do tema para professores e profissionais, bem como a especificidade dos conhecimentos necessários para o desempenho na área, tanto para uso como para elaboração de linguagens documentárias.

O levantamento bibliográfico realizado trouxe trabalhos sobre definição, métodos de elaboração e gestão de linguagens documentárias, e aportes teóricos e práticos para

ensino, formatação e uso dos sistemas de organização do conhecimento. Foram mostradas questões relacionadas às competências que precisam ser ensinadas, a interdisciplinaridade e tendências para desenvolvimento de sistemas e interoperabilidade que precisam ser incluídas nos cursos.

Devido à relevância das questões ligadas a ensino na área, consideramos importante a continuidade deste trabalho, analisando a literatura e experiências realizadas para propor formas de melhoria dos processos de ensino-aprendizagem e desenvolvimento de competências.

Referências

CAMPOS, M. L. A., GOMES, H. E. Metodologia de elaboração de tesauro conceitual: a categorização como princípio norteador. **Perspectivas em Ciência da Informação**, v.11, n.3, 348-359, 2006.

CARLAN, E.; BRÄSCHER, M. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **RICI:** Revista Ibero-americana de Ciência da Informação, v. 4, n. 2, p. 53-73, ago./dez. 2011.

CARO-CASTRO, Carmen. Léxico y documentación: del lenguaje natural al lenguaje documental. In: ROMAN, Maria Teresa Fuentes; REY, Jesus Torres del. Nuestras palabras: entre el léxico y la traducción. Frankfurt: Veuver, 2006. p. 127-145.

CLARKE, Stella G. Dextre; ZENG, Marcia Lei. From ISO 2788 to ISO 25964: the evolution of thesaurus standarts towards interoperability and data modeling. **ISQ** n. 1, v. 24, p. 20-26, 2012.

ESPELT, C. Bases teóricas em la enseñança de linguajes documentales. I Conferencia ISKO- Espanha, Madrid, 4-5 Novembre, 1993. **Proceedings**. Saragosa, 1995.

ESTEBAN-NAVARRO, M. A. Aplicaciones de la Terminologia para la docência de la gestión de lenguajes documentales. I Conferencia ISKO- Espanha, Madrid, 4-5 Novembre, 1993. **Proceedings**. Saragosa, 1995.

HEDDEN, H. Controlled vocabularies, thesauri and taxonomies. **The Indexer**, v. 26, n. 1, p. 33-34, 2008.

LEISE, F. Controlled vocabularies: an introduction. **The Indexer**, v. 26, n. 3, p. 121-126, 2008.

MAI, Jens-Erik. Actors, domains, and constraints in the design and construction of controlled vocabularies**. Knowledge organization**. v. 35, n. 1, p. 16-29, 2008.

MOREIRO-GONZÁLEZ, José Antonio. La enseñanza de los lenguajes documentales en las diplomaturas de Biblioteconomía y Documentación: problemática y tendencias. I Conferencia ISKO- Espanha, Madrid, 4-5 Novembre, 1993. **Proceedings**. Saragosa, 1995.

MORGAN, J.; BAWDEN, D. Teaching knowledge organization: educator, employer and professional association perspectives. **Journal of Information Science**, v. 32 p. 108, 2006.

NATIONAL INFORMATION STANDARDS ORGANIZATION. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**, 2010. (ANSI/NISO Z39.19-2005 R2010.)

PASTOR-SÁNCHEZ, J.; MARTÍNEZ-MÉNDEZ, F.;RODRÍGUEZ-MUÑOZ, J. Aplicación de SKOS para la interoperabilidad de vocabulários controlados en el entorno de linked open data. . **El profesional de la información**, v. 21, n. 3, p. 245-253, mayo-junio 2012.

Apoio: Universidade de São Paulo. Programa Ensinar com Pesquisa.

A genealogia do conceito de Ciência da Informação no Brasil: uma análise discursiva a partir de periódicos fundacionais na área

Larissa de Mello Lima
Universidade Estadual Paulista
larissalima.unesp@gmail.com

João Batista Ernesto de Moraes
Universidade Estadual Paulista
jota@marilia.unesp.br

Daniel Martínez-Ávila
Universidade Estadual Paulista
dmartinezavila@marilia.unesp.br

1 Introdução

O conceito de Ciência da Informação surgiu no Brasil, na decada de 1970, em um contexto universitario e como produto de práticas e transformações discursivas. Para apresentar a genealogia do conceito de Ciência da Informação no Brasil, torna-se fundamental realizar uma caracterização prévia dos agentes responsáveis por moldar o surgimento da área em contexto nacional.

O periódico *Ciência da Informação* é considerado um dos periódicos mais importantes da área e foi criado em 1972 pelo então IBBD, atual IBICT. No mesmo ano

o instituto incentivou que a Escola de Biblioteconomia da UFMG criasse a *Revista da Escola de Biblioteconomia da UFMG*. Essa iniciativa por parte do IBICT, é reflexo da criação do primeiro curso de mestrado em Ciência da Informação em 1970. (MENEZES; SILVA; COZINET, 2007).

É possível afirmar que a conjuntura histórica da Ciência da Informação no Brasil possui alicerces estruturados sob o viés institucional. Denomina-se por instâncias de validação científica da Ciência da Informação no Brasil as instituições que caracterizam e até mesmo determinam a história da área no país como o IBDD e os periódicos citados acima.

O delineamento de um discurso da área possui características evidentemente próprias ao discurso científico (ou o discurso que *busca* ser científico), com traços que tornam a análise pouco fluida, pela própria atuação da ideologia.

Pode-se compreender que o discurso científico, no qual a Ciência da Informação se alia, existe uma constante redefinição e repetição de posturas anteriores e de deslocamentos provocados pela ação do poder (instituições de pesquisa, pesquisadores, agências de fomento, universidades).

Pode-se dizer que é parte do próprio movimento de validação científica, ou seja, na medida em que um autor X de menor “impacto” cita autor Y de maior impacto e reconhecido pela comunidade científica (poder), o autor X busca filiar-se ao ponto de “superar” o autor Y, desta forma é possível dizer que a prática científica é refém de suas formulações enunciativas.

Existe, nos artigos analisados, um “jogo” que remete à ação ideológica e às instituições que legitimam a atuação desta ideologia, na medida em que o discurso é uma instância que registra essa atuação.

Ou seja, no âmbito da tessitura discursiva é importante situar que é sobre a atuação do sujeito que o discurso se constrói, mas tal atuação está condicionada à ação direta da ideologia que se traduz na multiplicidade de posições que o sujeito pode assumir no discurso recaindo assim no âmbito institucional.

2 Metodologia

A metodologia pauta-se nos pressupostos teóricos metodológicos da análise do discurso de matriz Francesa (FOUCAULT 1986; 2010, PÊCHEUX 1975)

A Análise do Discurso, no contexto da Ciência da Informação, já foi apresentada ou utilizada em diversos estudos teóricos e práticos (por exemplo, FROHMANN 1992; 1994a; FROHMANN 1994b; 2001; BUDD & RABER 1996; BUDD 2006; RADFORD 2003; RADFORD & RADFORD 2005; HAIDER & BAWDEN 2007; CAMPBELL 2007; 2011; OLSSON 2010; MARTÍNEZ-ÁVILA 2012).

Neste estudo serão analisados artigos selecionados dos periódicos *Ciência da Informação* e *Revista da Escola de Biblioteconomia da UFMG* da década de 1970, período de nascimento da Ciência da Informação no Brasil, na medida em que ambos se configuram como instâncias de validação científica da área em contexto nacional.

No que tange à seleção dos artigos é possível realizar algumas considerações que fornecem um esboço de algumas características inerentes ao discurso científico. O primeiro esboço que merece destaque é o fato de que o objeto de análise está inserido na tipologia do texto científico; este por sua vez é formado por comunidades científicas que são, também, comunidades discursivas na medida em que legitimam certas instâncias enunciativas em detrimento de outas.

No que tange ao método, a construção do dispositivo de observação traz alguns elementos que, quando articulados nos artigos durante a análise, delinearão o percurso conceitual e discursivo da disciplina. Situar esta análise enquanto arqueológica significa descrever como as práticas discursivas estão situadas em determinado ponto.

> Análise de discurso arqueológico não é intrinsecamente preocupada com o que deveria ser, no sentido de alcançar a resolução final para um enigma; Ela está preocupada com práticas discursivas como eles estão em um ponto no tempo. (BUDD, 2006, p. 73 tradução nossa).

Para descrevê-las é retomada a estrutura de Pêcheux (1975) revista por Orlandi (2000) e a esquematiza em etapas norteadas por perguntas fundamentais.

- “Primeiro tratamento de análise superficial”

Momento em que se tem um contato primário com a superfície linguística do texto. É também neste momento em que é exposto o elemento do arquivo, ou seja, o corpus que será submetido à análise.

- “Transformação da superfície linguística em objeto discursivo”
Para efetuar esta transformação é necessário realizar uma pergunta norteadora: “O que é dito neste discurso? O que é dito em outro discurso?” A partir de tal estruturação, expõe-se o objeto discursivo a partir dos fenômenos linguísticos discursivos (paráfrase, polissemia, polifonia) que incidem sobre ele.

- “Do objeto discursivo para o processo discursivo”:
Momento em que a pergunta norteadora é: “Por que isso e não outro?”. Na resposta em cada análise será atingido o processo discursivo, que mostra a relação que aquele dizer tem com o seu exterior.

3 Resultados

Na Tabela 1 mostram-se os conceitos de CI extraídos dos artigos selecionados dos periódicos *Ciência da Informação* e *Revista da Escola de Biblioteconomia da UFMG* na década de 1970.

Tabela 1: Quadro comparativo entre conceitos

ARTIGOS	CONCEITO DE CI
“Da Bibliografia à Ciência da Informação: Um Histórico e Uma Posição”	Disciplina científica que realiza investigações. (é o texto)
“Curso sobre ciência da informação para estudantes de biblioteconomia”	Ciência que refina os processos, sistemas e fênomenos da biblioteconomia.

“Relações Bibliométricas Entre a Frente de Pesquisa (Research Front) e Revisões da Literatura: Estudo Aplicado a Ciência da Informação”	. Disciplina que diz respeito aos fenômenos de comunicação
“Las ciencias de la información en la escuela de bibliotecología”	Ciência interdisciplinar que investiga as propriedades e comportamento da informação
“Distribuição da Literatura Geológica Brasileira: Estudo Bibliométrico”	Ciência baseada nos fundamentos da bibliometria.
“A pós graduação em Biblioteconomia.”	Ciência que vem auxiliar o ensino de Biblioteconomia
“Futurologia - Doença Infantil da Biblioteconomia”	Disciplina de caráter puramente científico.
“Ensino de Ciência da Informação: Bibliografia analítica.”	Ciência de caráter interdisciplinar
Comportamento bibliométrico da língua portuguesa, como veículo de representação da informação”	Ciência interdisciplinar na medida em que a relaciona, neste caso, com Linguística e Bibliometria
“Planejamento de uma unidade didática em Ciência da Informação.”	Ciência de caráter interdisciplinar
“Educação em Ciência da Informação na Década de 1980”	Ciência interdisciplinar que relaciona a prática profissional com a pesquisa científica.
“Pós graduação em biblioteconomia e Ciência da Informação: reflexões, sugestões, experiências”	Ciência de caráter basicamente interdisciplinar.

Fonte: Elaborado pelos autores.

Nota-se que, ao longo dos anos em que os artigos foram publicados, O IBBD, uma das instâncias de validação científica da Ciência da Informação no país, tinha como objetivo prioritário desenvolvimento em Ciência e Tecnologia para suprir uma lacuna no

sistema nacional de desenvolvimento científico e tecnológico conforme já ressaltado por Menezes; Silva; Couzinet (2007).

Acredita-se que haja por parte deste órgão uma aproximação com a linha de pensamento de Vannevar Bush que confere à Ciência da Informação o caráter de ciência que surge para solucionar a questão do aumento exponencial da produção científica do pós guerra. Ou seja, o IBICT passa a refletir, a partir de 1976, o que foi preconizado por Bush em 1945, que seria dar “atenção” à perspectiva tecnologica de processamento de dados.

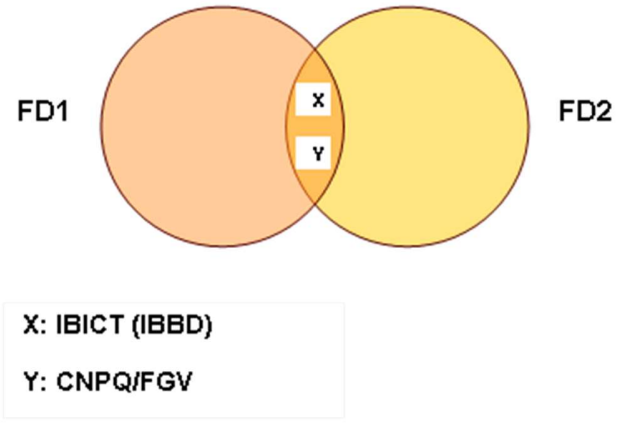
Em contraposição aos ideais teóricos do órgão em busca de efetiva valorização da perspectiva tecnológica, nota-se através do quadro comparativo, a explicitação de duas formações discursivas incidindo de maneira direta na área.

Os apontamentos de Smit, Tálamo e Kobashi (2004) convergem para a caracterização da formação discursiva 1, ilustrando assim que a questão da interdisciplinaridade é propriamente discursiva, na medida em que segue a formação de constituição da ciência embasada na pós modernidade sem conter de fato um norteamento claro sobre o que significa esta interdisciplaridade.

Como formação discursiva 1 (FD1) há a caracterização da Ciência da Informação a partir de sua interdisciplinaridade. Como formação discursiva 2 (FD2) há o entendimento de que a Ciência da Informação é uma disciplina recente que carece de balizas teóricas.

No ponto de intersecção entre ambas as formações discursivas (X;Y) temos o IBICT e o CNPq/FGV que se configuram como instâncias de validação científica da Ciência da Informação, na medida em que o IBICT foi o órgão responsável pela criação de ambos os periódicos que tiveram seus artigos analisados. E este órgão possui uma relação direta com o CNPq e a FGV por conta das iniciativas pioneiras destes que culminaram na sua criação. Ou seja, são fatores que dentro de uma conjuntura política formam a perspectiva de institucionalização da área.

Figura 2 - Formações discursivas da Ciência da Informação.



Fonte: Elaborada pelos autores

4 Conclusão

Conclui-se que a genealogia do conceito de Ciência da Informação na área conta com a presença de duas formações discursivas na década de 1970. Na FD1 há a perspectiva de interdisciplinaridade incidindo sobre ela. Na FD2 o fato da área ser recente carecendo de mais pesquisas científicas que consolidem suas balizas teóricas.

Entende-se sob a perspectiva deste estudo que a institucionalização histórico-discursiva da área no Brasil na década de 1970 forma a genealogia do conceito de Ciência da Informação no país se estruturando mediante formações discursivas que ora buscam sedimentar a questão da interdisciplinaridade que incide sobre a área, ora ressaltam o fato da Ciência da Informação ser recente, necessitando de balizas teóricas bem definidas.

Referências

BUDD, J. Discourse Analysis and the Study of Communication in LIS. **Library Trends**, v. 55, n. 1, p. 65-82, 2006.

BUDD, J.; RABER, D. Discourse Analysis: Method and Application in the Study of Information. **Information Processing & Management**, v. 32, n. 2, p. 217-226, 1996.

CAMPBELL, D. G. The Birth of the New Web: A Foucauldian Reading of the Semantic Web. **Cataloging & Classification Quarterly**, v. 43, n. 3-4, p. 9-20, 2007.

CAMPBELL, D. G. RDA and RDF: A Discourse Analysis of Two Standards of Resource Description. In: SMIRAGLIA, Richard P. (Ed.) **Proceedings from North American Symposium on Knowledge Organization, Vol. 3**, 2013, p. 207-216.

FOUCAULT, M. **A arqueologia do saber**. Rio de Janeiro: Forense, 1986.

FOUCAULT, M. **A ordem do discurso**. São Paulo: Edições Loyola, 2010.

FREITAS, L. S. **Na Teia dos Sentidos**: análise do discurso da Ciência da Informação sobre a atual condição da informação. 2001. Tese (Doutorado em Ciência da Informação e Documentação) - Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2001.

FROHMANN, B. The Power of Images: A Discourse Analysis of the Cognitive Viewpoint. **Journal of Documentation**, v. 48, n. 4, p. 365-386, 1992.

FROHMANN, B. Discourse Analysis as a Research Method in Library and Information Science. **Library and Information Science Research**, v. 16, p. 119-138, 1994a.

FROHMANN, B. Communication Technologies and the Politics of Postmodern Information Science. **Canadian Journal of Information and Library Science,** v. 19, n. 2, p. 1-22, 1994b.

FROHMANN, B. Discourse and Documentation: Some Implications for Pedagogy and Research. **Journal of Education for Library & Information Science** v. 42, n. 1, p. 13-26, 2001.

HAIDER, Jutta; BAWDEN, David. Conceptions of “information poverty” in LIS: a discourse analysis. **Journal of Documentation**, v. 63, n. 4, p. 534-557, 2007.

MARTÍNEZ-ÁVILA, D. Problems and Characteristics of Foucauldian Discourse Analysis as a Research Method. In: SMIRAGLIA, Richard P.; Lee, Hur-Li (Eds.), **Cultural Frames of Knowledge**. Würzburg: Ergon, 2012. p. 99-110.

MENEZES, E. M; SILVA, E. L; COUZINET, V. A Ciência da Informação na França e no Brasil. **Datagramazero**. v.8 , n. 6, 2007.

OLSSON, M. R. Michel Foucault: Discourse, Power/Knowledge, and the Battle for Truth.” In: LECKIE, Gloria J.; GIVEN, Lisa M.; BUSCHMAN, John E. (Eds.) **Critical**

theory for library and information science: exploring the social from across the disciplines. Santa Barbara: Libraries Unlimited, 2010, p. 63-74

ORLANDI, E. L. P. **Discurso e leitura**. 5. ed. São Paulo: Editora da Unicamp, 2000.

PÊCHEUX. M. **Semântica e discurso**: uma crítica à afirmação do óbvio. Tradução de Eni P. Orlandi [et al.]. Campinas: Editora da UNICAMP, 1975.

RADFORD, G. P. Trapped in our own discursive formations: toward an archaeology of library and information science. **The Library Quarterly**, v. 73, n. 1, p. 1-18, 2003.

RADFORD, G. P.; RADFORD, Marie L. Structuralism, post-structuralism, and the library: de Saussure and Foucault. **Journal of Documentation**, v. 61, n. 1, p. 60-78, 2005.

SMIT, J. W.; TÁLAMO, M. F.; KOBASHI, N. A determinação do campo científico da ciência da informação: uma abordagem terminológica. **Datagramazero***,* v. 5, n. 1, fev. 2004.

A Dimensão Aplicada da Organização e Representação do Conhecimento

A Discourse Analysis of the Semantic Web

D. Grant Campbell
University of Western Ontario
gcampbel@uwo.ca

This presentation will examine the Semantic Web as a discourse that rests on a set of unspoken political and social assumptions that distinguish it from traditional bibliographic control on the one hand, and from Big Data and social media systems on the other. Using Foucault's concept of the discursive formation, together with Charles Taylor's concept of the "social imaginary," this presentation will argue that the Semantic Web operates on a principle of shared responsibility and assumptions of honesty that preserve a link between the purpose of data collection and its ultimate use.

The Semantic Web has been around nearly as long as the Web itself, and has undergone numerous transformations in its conceptions and its implementations, the most recent being the rise of linked data. Tim Berners-Lee originally envisioned it as a ubiquitous network of data rather than of documents: a network that would mine the resources of the Deep Web—data contained in Web-accessible relational databases, and other structured contexts—to create systems that were capable of performing sophisticated inferences, checking for data quality and validity, and assembling data to respond to specific questions and needs (Berners-Lee, Hendler, &Lassila 2001). In its first imagined incarnations, the Semantic Web looked much like the data-saturated world in which we now live. Human beings went about their daily lives doing much of what they'd always done, but accompanied by intelligent agents that would relate ongoing activities to a data infrastructure, prompt human beings with suggestions, and proactively adjust the environment according to pre-established requirements.

This infrastructure would rest upon a series of standards established and maintained by the World Wide Web Consortium, including the Resource Description Framework (RDF), the RDF Schema, the Working Ontology Language (OWL), and Friend of a Friend (FOAF).Since the Consortium has no legislative or coercive function, no one forces an institution or a developer to use these standards. But since these standards address such ubiquitous processes as information description, classification, vocabulary control and social networking, the Consortium encourages organizations to opt in to these standards, thereby making the world gradually more interoperable, releasing data from its many silos to be combined and recombined for various purposes.

This utopian vision of a data infrastructure has two important features. First it is created by the actual users and creators of the data, rather than by a designated intermediary such as a library: users and creators who consciously decide to use Semantic Web standards. Second, it grows from a core of deliberate and purposeful statements. RDF is based on the concept of the triple—subject, predicate and object—in which a specific community of data creators and users expresses consensus about the attributes of a resource and its relationships with other resources. We have, therefore, a community of data creators, system designers and information users taking it upon themselves to make statements about data. As Michel Foucault reminds us, these statements, like all discursive formations, have power relationships embedded into them. By their very existence, they form images of the conditions under which statements emerge; the authorities that determine which statements can or cannot be made and by whom; and the grids of classifications that group and differentiate various statements (Foucault 2010).

These discursive formations enable us to clarify how Semantic Web practices differ from those of library systems on the one hand, and those of Big Data on the other. Libraries have traditionally assigned trained intermediaries, in the form of cataloguers and indexers, to create their own discursive formations in the form of catalogue records, classification systems, indexes and thesauri. End users have traditionally had to abide by the categories and descriptions created by these intermediaries.

Big Data systems, on the other hand, have a very different view of the surfaces of emergence. In the case of Big Data, what matters is not the statement, but the trace.

Uses of information typically leave traces that, when collected, aggregated and studied for patterns, enable organizations to derive insights and predict outcomes that are often unrelated to the original data use. Big data for targeted advertising and recommender systems might well study the time of hovering over a Netflix option, and the kinds of programs enjoyed on specific devices; Big Data for predictive analysis might use search queries, geographic location and social media behaviour. In these cases, what matters is not the statement—the purposeful articulation of data characteristics for the purposes of meeting specific needs—but the lingering trace of these statements, and the accumulation of these traces into something that bears only an incidental relationship to the data's original purpose.

Charles Taylor suggests that in our modern life and culture we are haunted by "social imaginaries": inherited assumptions and theories that were formulated long ago, and which now lie half buried and only partially recognized (Taylor 2005). If we take his concept and apply it to the discourse of the Semantic Web, compared to that of traditional libraries and to Big Data, we find an intriguing parallel with concepts of political economy that emerged in the seventeenth and eighteenth centuries. Libraries work on an assumption that owes much to Thomas Hobbes's *Leviathan*. Information requires a social contract, very similar to Hobbes's social contract, in which information users sign over their autonomy to a designated body: in so doing, they are saved from a chaotic and brutal state of nature (Hobbes 1985). Big Data works on a principle similar to that of Adam Smith's invisible hand: the notion that individual acts of self-interest, when aggregated together, achieve a social good that was unrelated to the selfish motivations of the individual acts (Smith 2000, 485).

The Semantic Web, however, echoes Jean-Jacques Rousseau's concept of the Social Contract: an arrangement whereby a community voluntarily takes control of itself, and by sacrificing freedom in a superficial sense achieves a more profound freedom (Rousseau 1988, 24). The creators of the Semantic Web envision an environment in which enlightened data users, data creators, and data system designers voluntarily choose to adopt a set of mutually agreed-upon standards at some immediate cost in autonomy, speed and cost, in order to create a stronger mutual empowerment within and between knowledge domains and information communities.

References

Berners-Lee, T., Hendler, J., &Lassila, O. (2001).The Semantic Web. *Scientific American* May 2001: 35-43.

Foucault, M. (2010).*The archaeology of knowledge and The discourse on knowledge*. New York: Vintage.

Hobbes, T. (1985).*Leviathan*. Middlesex: Penguin.

Rousseau, J-J. (1988). *On the social contract*. Indianapolis: Hackett.

Smith, A. (2000). *The wealth of nations*. New York: Modern Library.

Taylor, C. (2005). *Modern social imaginaries*. Durham: Duke University Press.

Tecnologias da web semântica aplicadas a organização do conhecimento: padrão SKOS para construção e uso de vocabulários controlados descentralizados

José Eduardo Santarem Segundo
Universidade de São Paulo
santarem@usp.br

Caio Saraiva Coneglian
Universidade Estadual Paulista
caio.coneglian@gmail.com

1 Introdução

A criação do projeto da Web Semântica, de Tim Berners-Lee, surgiu em face das dificuldades de localização, descrição e recuperação de informações em ambientes Web, portanto está diretamente relacionado a todo tipo de ciência, disciplina ou área que lida com estes processos.

A ideia da Web Semântica nasce a partir da necessidade de agentes computacionais interagirem com sistemas sem a necessidade de intervenção humana. Para isso é necessário o uso de uma grande quantidade de tecnologias e conceitos.

Caracteriza-se a Web Semântica como um projeto que visa dispor nos sites tanto informações descritivas e temáticas para os usuários, como informações que possam ser processadas e identificadas pelos computadores automaticamente. Assim, seria uma forma de disponibilizar informações para as máquinas/softwares juntamente com as informações para os usuários (BERNERS-LEE, LASSILA, HENDLER, 2001).

A Web Semântica dá estrutura ao significado da página Web, criando um ambiente propício para que os agentes de busca possam realizar tarefas sofisticadas e entregá-las ao usuário (BERNERS-LEE, LASSILA, HENDLER, 2001).

O desafio da Web Semântica vem sendo, a cada dia, prover uma linguagem capaz de expressar ao mesmo tempo dados e regras, de forma a possibilitar a dedução de novos dados e regras a partir de qualquer sistema de representação de conhecimento a ser importado ou exportado na Web.

As tecnologias e conceitos que envolvem a Web Semântica tem como ponto fundamental a criação de uma nova estrutura de armazenamento de dados. O ponto principal está na separação da apresentação e da estrutura do conteúdo disponível, tratando as unidades atômicas de uma informação como componentes independentes.

Essa separação permitirá uma recuperação da informação de várias maneiras, independente de como seja a busca, bastando que se conheça a estrutura dos dados. Este novo formato de recuperação de informação deverá facilitar a associação entre informações e ajudará a minimizar o problema da utilização de uma mesma informação em vários sistemas.

Apesar da Web Semântica ser apresentada como uma estrutura de sete camadas, o passar dos anos vem nos mostrando que as tecnologias e linguagens que compõe estas camadas podem ser aplicadas individualmente ou em agrupamentos, de forma a construir ambientes digitais que permitam armazenamento e recuperação da informação com significado, ou seja, utilizando-se de relações semânticas.

No contexto da organização da informação, entende-se que essas tecnologias podem ser utilizadas na construção e uso de vocabulários controlados, de forma a facilitar a descentralização e automatização do uso dos mesmos.

Figura-se como problema nesta pesquisa a dificuldade que as organizações tem de transformar as linguagens documentárias em objetos digitais independentes, passíveis de uso de forma unificada e integrada, por vários tipos de plataformas digitais. Os principais exemplos de plataformas que necessitam de vocabulários controlados são os catálogos de bibliotecas, os repositórios digitais e as revistas científicas eletrônicas. Há ainda uma grande quantidade de outras plataformas e serviços de informação que poderiam utilizar vocabulários controlados.

A necessidade de uso constante de linguagens documentarias nesses ambientes é latente e os processos de uso ainda são rudimentares, exigindo procedimentos manuais para que o processo possa ser executado a contento.

Destaca-se ainda neste trabalho, que consideramos procedimento manual o ato do responsável pela indexação ter de realizar o processo de uso do vocabulário controlado acessando ambientes externos a plataforma em uso (repositórios, catálogos ...), ou seja, recorrendo a sites, sistemas proprietários, documentação em planilhas e outros tipos de recurso digital. Ressalta-se ainda que esta pesquisa em momento algum trata de indexação automática, ficando esse estudo restrito a integração de linguagens documentarias a plataformas digitais de uso contínuo.

O objetivo principal deste trabalho é apresentar o uso da ontologia SKOS, e de melhores práticas de publicação como o Linked Data, como artefato computacional necessário a construção de vocabulários controlados para uso distribuído em ambientes descentralizados.

Apresentam-se ainda como objetivos específicos a apresentação da ontologia SKOS, os conceitos básicos de Linked Data e também a sugestão de uso de vocabulários controlados por plataformas digitais.

2 SKOS (Simple Knowledge Organization System)

SKOS é um modelo padronizado de publicação de sistemas de organização do conhecimento na Web, baseado nos conceitos estruturais de ontologias da Web Semântica.

Apesar dos vocabulários controlados, e outros tipos de linguagens documentarias desta natureza, não compreenderem o uso de axiomas, inferências e instâncias, como o fazem as ontologias, o modelo SKOS foi concebido de forma que tecnicamente, ou tecnologicamente, uma linguagem documentaria escrita usando este padrão seja representada e principalmente factível de consulta e recuperação de dados, como as ontologias utilizadas nos padrões da Web Semântica.

Implica-se, portanto, que essas estruturas de representação do conhecimento podem ser ampliadas em suas possibilidades, com o uso de relações que são utilizadas na construção de ontologias.

Há uma porção de definições para os conceitos de ontologia, alguns deles ligados a Filosofia ou associados a ideia de taxonomia.

De acordo com Santarem Segundo e Vidotti, (2011, p. 287),

> Ontologias fornecem o conhecimento estruturado e uma infraestrutura para integrar bases de conhecimentos, independentes da implementação. Constituem uma ferramenta poderosa para suportar a especificação e a implementação de sistemas computacionais de qualquer complexidade.

Para o uso como tecnologia da Web Semântica, entende-se as ontologias como: artefatos computacionais que descrevem um domínio do conhecimento de forma estruturada, através de: classes, propriedades, relações, restrições, axiomas e instâncias.

O SKOS tem sua especificação, baseada em linguagem OWL, como uma ontologia completa, que pode ser expresso usando um modelo de representação do conhecimento através do Resource Description Framework (RDF). Usar RDF permite que ele seja transmitido entre aplicações informáticas de forma interoperável. (SKOS, 2004).

Segundo o W3C, o RDF é uma linguagem de uso geral para representar informações na Web. O RDF tem como princípio fornecer interoperabilidade aos dados, de forma que possa contribuir com a recuperação de informações de recursos na Web.

O modelo RDF é constituído de três objetos básicos: recursos, propriedades e declarações. Um recurso é uma informação que pode ser identificada por uma URI (Universal Resource Identificator). Propriedades são as informações que representam as características do recurso, ou seja, são os atributos que permitem distinguir um recurso de outro ou que descrevem o relacionamento entre recursos. A declaração é a constituição da informação completa, que compreende um recurso com suas propriedades e valores para as propriedades.

O modelo SKOS divide-se em 9 grupos: conceitos de classe, conceitos de esquema, rótulos, notações, propriedades, relações semânticas, coleções e mapeamento de propriedades. Esses grupos são representados por elementos (como: skos:prefLabel, skos:altLabel, entre outros) nomeados vocabulários, que efetivam a relação existente entre os termos em um vocabulário controlado.

3 Linked Data

Pensando em um modelo associativo de publicação de dados estruturados na Web, foi constituído o *Linked Data* (dados ligados). Organizado por Tim Berners-Lee tem como característica principal o estabelecimento de links entre dados de fontes distribuídas.

Segundo Heath e Bizer (2011), “o Linked Data é um conjunto de melhores práticas para publicação e conexão de dados estruturados na Web, permitindo estabelecer links entre itens de diferentes fontes de dados para formar um único espaço de dados global”.

Para Berners-Lee (2006),

> a Web Semântica não trata apenas de depósito de dados na web. Trata-se de fazer ligações, de modo que uma pessoa ou máquina possa explorar esse conjunto de dados. Com *Linked Data*, quando você tem um pouco de dados, você pode encontrar outros que estão relacionados.

O projeto tem crescido muito nos últimos anos, em 2007 o *Linked Data* era constituído de aproximadamente um bilhão de declarações RDF, interligados por 120.000 links RDF, constituídos em aproximadamente 25 datasets. Em 2011 eram 52 bilhões de declarações RDF, em 295 datasets. Atualmente são 570 datasets.

A base de maior tamanho e também maior destaque no *Linked Data* é a *DBpedia*. A principal referência de instrumentos de controle de vocabulário e controle de autoridade no Linked Data é o *AGROVOC,* que é um tesauro desenvolvido a partir de 1980 que atualmente é referência para temas ligados à agricultura, piscicultura, silvicultura entre outros assuntos relacionados ao meio ambiente.

O *AGROVOC* é utilizado mundialmente por pesquisadores, bibliotecários, gestores de informação e outros profissionais para indexar, recuperar e organizar dados em sistemas de informações agrícolas. Utilizando-se da terminologia *SKOS*, modelo apropriado para publicação de tesauros na web, o *AGROVOC* pode ser consultado assim como interligado com outros recursos para provimento de dados em formato semântico. (AGROVOC, 2014).

Usando o modelo SKOS e também as melhores práticas do Linked Data, é possível construir e disponibilizar para a comunidade, de forma aberta, qualquer vocabulário controlado.

4 Construção e uso de vocabulários controlados de forma descentralizada

Conforme verificou-se, atualmente dispomos de um conjunto de tecnologias e ferramentas que permitem a construção e a publicação em formato aberto de um vocabulário controlado usando SKOS. Utilizar-se das melhores práticas do Linked Data é um dos caminhos para isso.

Ao analisar os vocabulários controlados de grande parte das instituições brasileiras, avaliando a manutenção e principalmente seu uso, chegamos ao problema que norteia esta pesquisa, a dificuldade em disponibilizar de forma descentralizada os vocabulários controlados.

Algumas organizações utilizam-se de sistemas que permitem a manutenção automatizada do vocabulário controlado, visto que, há no mercado sistemas de automação de bibliotecas que dispõe de módulos para essa finalidade. Entretanto identificam-se organizações que ainda usam ferramentas ultrapassadas ou técnicas pouco usuais para construir e dar manutenção em seus vocabulários controlados.

O uso dos vocabulários controlados, pelas equipes de indexação em suas plataformas digitais, é ainda mais complicado que o processo de construção e manutenção destes. Verificam-se diferentes processos e procedimentos para utilização dos vocabulários controlados nas diversas plataformas que as unidades de informação tem de lidar em seu cotidiano.

Nos últimos anos além dos já conhecidos catálogos online (OPAC's), outras ferramentas e plataformas passaram a fazer parte do dia a dia das organizações e também necessitam do uso de vocabulários controlados para organização da informação. Revistas científicas em formatos eletrônicos, repositórios digitais, plataformas de integração entre bibliotecas e editoras para acesso a ebooks, gerenciadores de documentos eletrônicos,

entre outros, precisam da integração de vocabulários controlados para melhor organização e indexação de seu conteúdo.

As equipes de tratamento da informação, ainda não encontram nessas plataformas citadas, disponibilidade do vocabulário controlado para a indexação dos conteúdos. Em grande parte das vezes o procedimento é realizado de forma manual e totalmente desintegrada. Na maioria das vezes opta-se pelo acesso e consulta ao vocabulário controlado em um sistema e a inserção dos dados manualmente em outro sistema ou plataforma.

Esse procedimento pode gerar uma grande quantidade de problemas, desde os mais comuns, como a digitação indevida de termos recorrentes até as dificuldades com a atualização periódica dos termos, que gera uma árdua tarefa de atualização dos descritores nas plataformas que não estão integradas a linguagem documentaria.

Identifica-se ainda um outro problema quando existe a necessidade de realizar a exportação de um vocabulário controlado construído por uma instituição para que possa ser utilizado por outra. Nestes casos notam-se dificuldades com tecnologias que permitam carregar a estrutura do vocabulário controlado. Essa dificuldade muitas vezes é ocasionada porque a linguagem documentária foi construída em uma ferramenta proprietária ou desenvolvida unicamente para uma determinada instituição, não dispondo de uma equipe técnica responsável pela atualização ou manutenção da ferramenta.

Quando existe a possibilidade de exportação de um vocabulário ele passa a ter vida própria na instituição que o está recebendo, sem contato com a instituição criadora, não recebendo mais as atualizações e manutenções que o original continua a receber pela equipe da instituição de origem.

Dado o contexto atual, as dificuldade de uso e manutenção dos vocabulários controlados e todas as dificuldades que a falta de integração causa entende-se que há uma necessidade latente de organização de uma estrutura que permite a integração direta de vocabulários controlados a vários tipos de plataformas, independente da ferramenta ou sistema que esteja sendo utilizado.

A proposta apresentada aqui caracteriza-se pelo uso de uma estrutura de vocabulário controlado que esteja sedimentada sobre tecnologias que permitam a

interoperabilidade entre sistemas e que o vocabulário controlado esteja disponível em um ambiente aberto para livre acesso por todo e qualquer sistema ou plataforma digital.

Entende-se que o uso das tecnologias da Web Semânticas, baseadas em RDF e organizadas pelo padrão SKOS, permitam a constituição de uma estrutura interoperável e passível dos dados serem atualizados e recuperados por máquinas, através da integração de sistemas.

A proposta, como pode ser observado na Figura 1, parte do princípio que uma instituição, ou um grupo de instituições, chamadas produtoras do vocabulário controlado, executem as tarefas de manutenção do conteúdo, enquanto outras instituições, nomeadas consumidoras, possam integrar o produto a todas suas plataformas e ferramentas para o uso diário.

A instituição, ou grupo de instituições nomeadas produtoras, também serão consumidoras do vocabulário controlado.

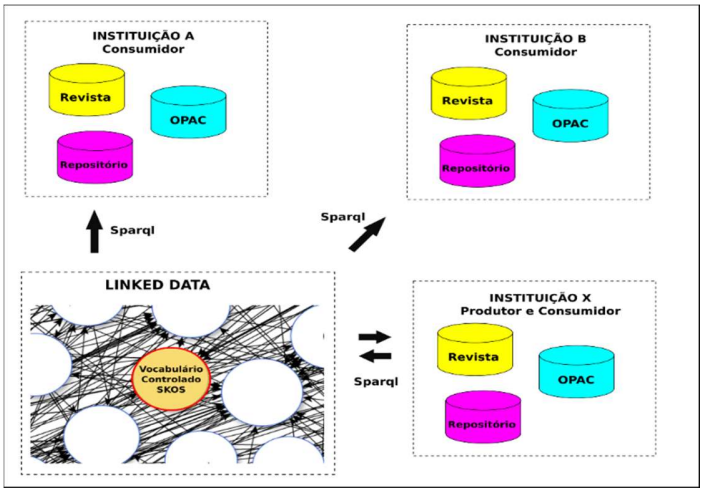


Figura 1 - Modelo de descentralização e uso integrado de vocabulário controlado.

Fonte: próprio autor

O produto da linguagem documentaria, baseado em SKOS, será publicado utilizando-se as melhores práticas do Linked Data e dessa forma estará disponível para uso aberto e livre de restrições, com aplicação de uma licença adequada.

Para que seja realizada a integração do vocabulário controlado, baseado em SKOS, será necessária a criação de plug-ins (partes de código) que possam ser integradas aos

mais variados tipos de sistemas já citados anteriormente, entretanto pelo vocabulário controlado ser construído com tecnologias abertas, a integração passa a plenamente possível.

O protocolo Sparql, será a principal tecnologia para recuperar e atualizar os dados pelas instituições produtoras e consumidoras do vocabulário controlado. Sparql é o nome dado a um conjunto de especificações que fornecem linguagens e protocolos para consultar e manipular o conteúdo publicado em RDF na Web. Atualmente chamado de protocolo, a versão 1.1 é uma recomendação W3C desde março de 2013. (SANTAREM SEGUNDO, 2014).

5 Considerações finais

Baseado nos objetivos propostos, esta pesquisa apresenta uma proposta de modelo de construção e uso de vocabulários controlados baseados em tecnologias e conceitos da Web Semântica.

Caracteriza-se o uso das linguagens RDF, Sparql e principalmente do modelo SKOS como propulsores ao uso de vocabulários controlados descentralizados, e livres de sistemas proprietários.

A solução baseada em publicação do vocabulário controlado utilizando-se das melhores práticas de publicação de dados abertos, baseados no Linked Data, representa uma novo paradigma de construção e uso de linguagens documentarias por sistemas e instituições diferentes, de forma colaborativa. Esse novo modelo, tem como princípio ser descentralizado e aberto, de forma que possa ser integrado a vários tipos de ferramentas e plataformas.

Conclui-se que o modelo apresentado, e o uso de tecnologias da Web Semântica, representam um novo conceito para a construção e uso de vocabulários controlados.

Referências

AGROVOC. AGROVOC Linked Open Data. 2014. Disponível em <http://aims.fao.org/standards/agrovoc/linked-open-data>.

BERNERS-LEE T.; LASSILA, O.; HENDLER, J. The semantic web. Scientific American, New York, v. 5, May 2001.

BERNERS-LEE, T. Linked Data Principles. 2006. Disponível em <http://www.w3.org/DesignIssues/LinkedData.html>.

HEATH, T.; BIZER, C. (2011). Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool, 1st edition.

SANTAREM SEGUNDO, J. E. Web Semântica: introdução a recuperação de dados usando SPARQL. Enancib, v. 15, 2014.

SANTAREM SEGUNDO, J. E.; VIDOTTI, S. A. B. G. Representação iterativa e folksonomia assistida para repositórios digitais. Liinc em Revista, v. 7, n. 1, p. 283-300, 2011. Disponível em: < http://dx.doi.org/10.18225/liinc.v7i1.414>.

SKOS. Simple Knowledge Organization System. Disponível em: <http://www.w3.org/2004/02/skos/>.

Análise preliminar sobre a conversão de plano de classificação em vocabulário controlado

Luciana Davanzo
Universidade Estadual Paulista
luciana.davanzo@gmail.com

Walter Moreira
Universidade Estadual Paulista
walter.moreira@marilia.unesp.br

1 Introdução

A classificação é uma das operações mais importantes do trabalho arquivístico, pois alicerça, de certo modo, quase todas as outras operações. Por meio dela são gerados dois importantes instrumentos arquivísticos: a) o plano de classificação dos documentos, que tem por objetivo representar as informações que foram produzidas e recebidas pelas instituições em função de suas atividades e b) a tabela de temporalidade documental, que têm por objetivo estabelecer os prazos de guarda dos documentos de arquivo.

O objetivo da classificação é, basicamente, “dar visibilidade às funções e às atividades do organismo produtor do arquivo, deixando claras as ligações entre os documentos, conforme salienta Gonçalves” (1998, p. 12).

Pode-se perceber a relevância da classificação arquivística, a partir da publicação, pelo Conselho Nacional de Arquivos (Conarq), vinculado ao Arquivo Nacional, da “Classificação, temporalidade e destinação de documentos de arquivo relativos as atividades-meio da administração pública”, no ano de 2001. Esse documento é parte do processo de “racionalização e controle eficazes das informações” (ARQUIVO NACIONAL, 2001, p. 7) contidas nos órgãos da administração pública.

A classificação arquivística é um dos pilares da gestão de documentos eficiente e eficaz e possibilita, por meio do ordenamento das relações entre os assuntos dos documentos, maior precisão tanto em relação ao processo de recuperação da informação quanto em relação ao controle das informações que fazem parte de uma instituição. Assim, todos os tipos de documentos, que são produzidos e recebidos pelas instituições, independente de seus suportes podem e devem ser devidamente classificados para que posteriormente sejam recuperados.

A eficácia na recuperação da informação, portanto, é condicionada à qualidade nos processos de representação da informação, como é, aliás, amplamente sabido. Sem representação clara e objetiva em relação aos interesses institucionais não poderá haver recuperação plenamente satisfatória pelo usuário.

Considerando-se, então, o valor e a importância da representação, tanto a representação documentária, realizada pelo profissional, quanto a representação da informação realizada pelo usuário no momento em que realiza a busca, acredita-se que os vocabulários controlados, elaborados com base nos planos de classificação arquivística, podem atuar como mediadores eficazes nas negociações de sentido que realizam sistema e usuário, mediadas pela linguagem em unidades de informação arquivística.

Tendo-se esse quadro em vista, propôs-se como objetivo para essa pesquisa realizar um experimento de “criação” de vocabulário controlado a partir de um plano de classificação estabelecido.

Embora o plano de classificação atue de forma positiva em relação ao acesso aos documentos, entende-se que, para uma recuperação mais precisa da informação, é necessário fazer o uso dos vocabulários controlados juntamente com os planos de classificação, afinal os vocabulários controlados irão garantir que os documentos sejam descritos a partir de padronização terminológica, o que contribuirá com a consistência em relação a sua representação.

O desenvolvimento da pesquisa justifica-se pela contribuição do uso dos vocabulários controlados pela arquivística porque esses são instrumentos que colaboram para o processo de organização e recuperação da informação. Os vocabulários controlados ainda são relativamente pouco utilizados pela arquivística, porque na maioria

das vezes ainda são entendidos como instrumentos auxiliares da biblioteconomia, área na qual são mais estudados e aplicados.

Dessa maneira, acredita-se que os instrumentos de controle e padronização terminológica agregam-se de forma positiva ao processo de gestão de documentos, pois estimulam o "cuidado em relação ao controle de vocabulário ao se nomear, por exemplo, as funções, atividades e séries documentais no contexto dos instrumentos de gestão documental [...]" (SMIT; KOBASHI, 2003).

A análise toma como base para a comparação e criação de vocabulário controlado, a subclasse “020 Administração de pessoal”, do plano de classificação proposto para atividades-meio da administração pública pelo Conarq (ARQUIVO NACIONAL, 2001). Esse plano de classificação possui dez classes principais, as quais “correspondem às grandes funções desempenhadas pelo órgão”. Seguindo-se a base decimal hierárquica, cada classe é subdividida sequencialmente em subclasses, grupos e subgrupos.

A Figura 1 apresenta uma representação esquemática da classe 000 (Administração geral) do plano de classificação. Na sequência, Figura 2, apresenta-se o primeiro nível da subclasse 020 (Administração de pessoal), o qual foi utilizado para compor o vocabulário controlado. A escolha dessa subclasse se deu, principalmente, em função dos limites do trabalho que ora se apresenta.

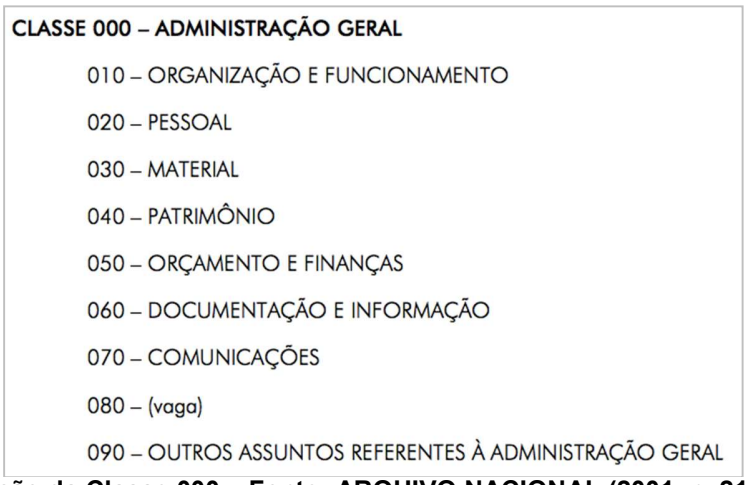
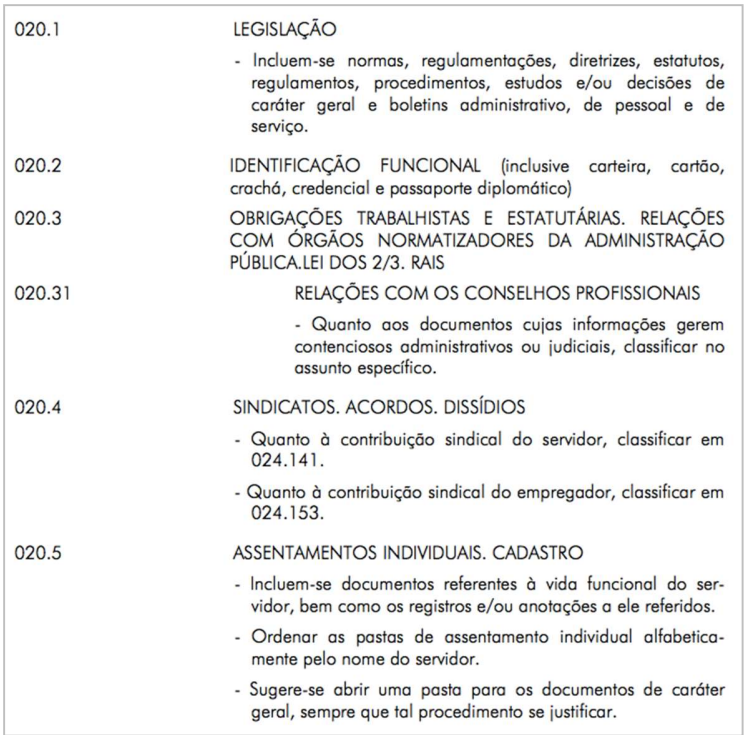

CLASSE 000 – ADMINISTRAÇÃO GERAL
010 – ORGANIZAÇÃO E FUNCIONAMENTO
020 – PESSOAL
030 – MATERIAL
040 – PATRIMÔNIO
050 – ORÇAMENTO E FINANÇAS
060 – DOCUMENTAÇÃO E INFORMAÇÃO
070 – COMUNICAÇÕES
080 – (vaga)
090 – OUTROS ASSUNTOS REFERENTES À ADMINISTRAÇÃO GERAL

Figura 2 - Organização da Classe 000 – Fonte: ARQUIVO NACIONAL (2001, p. 21)



020.1	LEGISLAÇÃO - Incluem-se normas, regulamentações, diretrizes, estatutos, regulamentos, procedimentos, estudos e/ou decisões de caráter geral e boletins administrativo, de pessoal e de serviço.
020.2	IDENTIFICAÇÃO FUNCIONAL (inclusive carteira, cartão, crachá, credencial e passaporte diplomático)
020.3	OBRIGAÇÕES TRABALHISTAS E ESTATUTÁRIAS. RELAÇÕES COM ÓRGÃOS NORMATIZADORES DA ADMINISTRAÇÃO PÚBLICA.LEI DOS 2/3. RAIS
020.31	RELAÇÕES COM OS CONSELHOS PROFISSIONAIS - Quanto aos documentos cujas informações gerem contenciosos administrativos ou judiciais, classificar no assunto específico.
020.4	SINDICATOS. ACORDOS. DISSÍDIOS - Quanto à contribuição sindical do servidor, classificar em 024.141. - Quanto à contribuição sindical do empregador, classificar em 024.153.
020.5	ASSENTAMENTOS INDIVIDUAIS. CADASTRO - Incluem-se documentos referentes à vida funcional do servidor, bem como os registros e/ou anotações a ele referidos. - Ordenar as pastas de assentamento individual alfabeticamente pelo nome do servidor. - Sugere-se abrir uma pasta para os documentos de caráter geral, sempre que tal procedimento se justificar.

Figura 3 – Organização da subclasse 020 – Fonte: ARQUIVO NACIONAL (2001, p. 23)

Como recurso técnico para essa pesquisa, que tem caráter exploratório, foi utilizado um *software* para a criação de tesauros denominado *Thew*, em versão Java, criado por Timothy C. Craven.

2 Vocabulários controlados e classificação arquivística

Esclarecendo sobre a operação de classificação e de ordenação, que lhe é subsequente, Heredia Herrera (1995, p. 263, tradução livre), define classificação como a ação de “separar ou dividir um conjunto de elementos estabelecendo classes ou grupos” e ordenação como “unir todos os elementos de cada grupo seguindo uma unidade de ordem, que pode ser data, ou o alfabeto, ou o tamanho ou um número”. Araújo (2006), discorrendo sobre a cotidianidade da classificação, define-a como um processo de “dividir em grupos ou classes, segundo as diferenças e semelhanças”.

Sobre esse olhar Costa (1998 *apud* ARAÚJO, 2006, p. 118) observa que “a ação de classificar é parte constitutiva das sociedades, na medida em que estão por toda parte, impregnam a vida social de forma onipresente”, dessa forma, o ato de classificar

acompanha toda a trajetória do homem, sendo então necessária, para o seu desenvolvimento em relação a sua visão de mundo.

A atividade de classificação, mais do que uma atividade meramente técnica, deve ser compreendida no escopo mais amplo das atividades de organização e representação da informação. A organização e representação do conhecimento é uma disciplina científico-aplicada

> cujo objetivo é melhorar a circulação da informação dentro e por meio de sistemas mediadores – os centros de seleção, armazenamento, recuperação e disseminação de informação – para produzir novo conhecimento e/ou facilitar o acesso ao [conhecimento] existente (GARCÍA MARCO, 1993, p. 100, tradução livre).

Um dos instrumentos utilizados pela organização e representação da informação para potencializar a circulação da informação é o vocabulário controlado. Os vocabulários controlados permitem a diminuição do uso de vários termos para representar uma única informação ou conceito, assim como contribuem para a diminuição de ruídos comunicacionais que podem existir entre usuários e sistemas de informação.

Os vocabulários controlados são ferramentais que corroboram tanto para o processo de organização da informação, como para o processo de recuperação da informação, por isso, se forem utilizados nos planos e/ou códigos de classificação, podem ser úteis em relação à polissemia de termos que podem existir no momento da indexação e representação da informação.

Portanto, os vocabulários controlados atuam como linguagem documentária e são construídos especialmente para “padronizar e facilitar a entrada e a saída de dados em um sistema de informações” (KOBASHI, 2008, p. 1); dessa forma, acredita-se que os vocabulários controlados diminuem os problemas oriundos da recuperação da informação porque estes atuam diretamente no controle da padronização terminológica.

O controle e a padronização terminológica devem ser observados como oportunidades para que haja melhorias no processo de recuperação, compartilhamento e socialização da informação, pois, “uma das principais dificuldades que os arquivistas encontram não é a multiplicidade de termos utilizados para designar os níveis de classificação, mas a falta de consenso sobre o significado desses termos”, conforme argumenta Sousa (2006, p. 7).

Principalmente no que diz respeito aos usuários não especializados, a utilização do vocabulário controlado torna-se mais do que necessária, indispensável, uma vez que estes usuários poderão realizar suas pesquisas de maneira mais profícua, assim como será evitado também a dispersão informacional, o que é uma risco bastante comum quando não existe um instrumento de controle vocabular.

Por esse motivo, acredita-se que a classificação é dos grandes pilares da arquivística. Os instrumentos originados dela vão servir de alicerce para os demais afazeres arquivísticos, por isso, entende-se que existe a necessidade da utilização de outros instrumentos que contribuam em relação a tarefa de representação e recuperação da informação, principalmente em tempos de avanços das tecnologias, afinal, a "classificação não pode responder mais sozinha pela importante e atualizada tarefa de busca à informação contida nos documentos de arquivo" (SOUSA, 2014, p. 21).

3 Resultados e discussão

A análise desenvolvida nessa seção toma por base a organização da subclasse 020 do Plano de Classificação do Conarq (ARQUIVO NACIONAL, 2001), conforme apresentado na Figura 2. Com base nessa subclasse e com o recurso do *software* Thew, conforme descrito anteriormente, realizou-se a conversão dos termos tratados nessa subseção para a estrutura de um vocabulário controlado. Na medida do possível, mantiveram-se os mesmos termos utilizados no Plano de Classificação, quando necessário – para preservar a estrutura do vocabulário controlado ou para exemplificação – foram inseridos novos termos nas relações hierárquicas, associativas ou de equivalência.

O Quadro 1 apresenta o resultado da conversão no formato sistemático do vocabulário controlado. Nesse modo de apresentação, não são mostrados os indicadores de relações entre os termos (TG, TE, TR, USE, UP) e as notas de escopo (SN). A escolha desse modo de apresentação se deu em função dos limites de extensão do trabalho fixados pela organização do evento.

- ADMINISTRAÇÃO DE PESSOAL
  - ACORDO COLETIVO DE TRABALHO
  - ASSENTAMENTO INDIVIDUAL

- CONSELHOS PROFISSIONAIS
  - CONSELHO FEDERAL DE ADMINISTRAÇÃO
  - CONSELHO FEDERAL DE BIBLIOTECONOMIA
  - CONSELHO FEDERAL DE MEDICINA
- CONVENÇÃO COLETIVA DE TRABALHO
- DISSIDIO
- IDENTIFICAÇÃO FUNCIONAL
  - CADASTRO FUNCIONAL
  - CARTÃO DE IDENTIFICAÇÃO
  - CARTEIRA DE TRABALHO E PREVIDÊNCIA SOCIAL
  - CRACHÁ
  - CREDENCIAL
  - PASSAPORTE DIPLOMÁTICO
- LEGISLAÇÃO
  - BOLETIM ADMINISTRATIVO
    - BOLETIM ADMINISTRATIVO DE PESSOAL
    - BOLETIM ADMINISTRATIVO DE SERVIÇO
  - DECISÃO DE CARÁTER GERAL
  - DIRETRIZ
  - ESTATUTO
  - ESTUDO DE CARÁTER GERAL
  - NORMA
  - PROCEDIMENTO
  - REGULAMENTO
- OBRIGAÇÕES TRABALHISTAS E ESTATUTÁRIAS
  - LEI DOS DOIS TERÇOS
  - RELAÇÃO ANUAL DE INFORMAÇÕES SOCIAIS
- ÓRGÃOS NORMATIZADORES DA ADMINISTRAÇÃO PUBLICA
- SINDICATO
  - CONTRIBUIÇÃO SINDICAL
    - CONTRIBUIÇÃO SINDICAL DO EMPREGADOR
    - CONTRIBUIÇÃO SINDICAL DO SERVIDOR

Quadro 1 – Apresentação sistemática do vocabulário controlado (Fonte: dados da pesquisa)

Diante da rede de relações conceituais estabelecidas entre os termos que foram selecionados do plano de classificação do Arquivo Nacional, pode-se notar que o plano de classificação, a partir do uso do vocabulário controlado torna-se mais claro em relação a representação da informação.

4 Considerações finais

Independentemente da instituição e do tipo de suporte, é necessário que todos os documentos sejam classificados porque “só o plano de classificação pode lhes conferir o sentido necessário de seu uso” (LOPES, 1996, p. 95). Dessa forma, salientamos que a partir da prática de controle de vocabulário será possível atender de modo mais eficaz às necessidades informacionais dos usuários, porque a partir da padronização dos termos utilizados para representar os documentos será possível a criação de um plano de

classificação a partir de uma linguagem clara, de acordo com a terminologia das áreas de domínio cobertas pelo arquivo.

Em estudos futuros, pretende-se ampliar a amostra e estender o experimento de conversão para outras classes do plano de classificação adotado como modelo.

Referências

ARAÚJO, C.A.A. Fundamentos teóricos da classificação. **Encontros Bibli**, n. 22, p. 117-140, jul./dez. 2006.

ARQUIVO NACIONAL (BRASIL). Conselho Nacional de Arquivos. **Classificação, temporalidade e destinação de documentos de arquivo relativos às atividades-meio da administração pública**. Rio de Janeiro: Arquivo Nacional, 2001.

GARCÍA MARCO, F. J. Paradigmas científicos en representación y recuperación de la información. In: GARCÍA-MARCO, Francisco Javier (Ed.). **Organización del conocimiento en sistemas de información y documentación**: actas del I Encuentro de ISKO-España, Madri. Zaragoza: Universidad de Zaragoza, 1993.

GONÇALVES, J. **Como classificar e ordenar documentos de arquivo**. São Paulo: Arquivo do Estado, 1998.

HEREDIA HERRERA, A. **Archivística General**: teoría y práctica. 7. Ed. Sevilla: Servicio de Publicaciones de la Diputación de Sevilla, 1995.

KOBASHI, N.Y. **Vocabulário controlado**: estrutura e utilização. Brasília, Escola Nacional de Administração Pública, 2008.

LOPES, L.C. **A informação e os arquivos**: teorias e práticas. Niterói: Eduff. 1996.

SMIT, J. W; KOBASHI, N. Y. **Como elaborar vocabulário controlado para a aplicação em arquivos**. São Paulo: Arquivo do Estado; Imprensa Oficial, 2003.

SOUSA, R.T.B. Alguns apontamentos sobre a classificação de documentos de arquivo. Brazilian Journal of Information Science: research trends, v. 8, n. 1, 2014.

_____. Classificação de documentos arquivísticos: trajetória de um conceito. **Arquivística.net**, v. 2, n. 2, p. 120-142, ago./dez. 2006.

Reengenharia de Tesauro: caso do Thesagro

Benildes C. M. S. Maculan
Universidade Federal de Minas Gerais
benildes@gmail.com

Gercina A. B. O. Lima
Universidade Federal de Minas Gerais
limagercina@gmail.com

Ivo Pierozzi Jr.
Embrapa Informática Agropecuária
ivo.pierozzi@embrapa.br

Leandro H. M. Oliveira
Embrapa Informática Agropecuária
ivo.pierozzi@embrapa.br

1 Introdução

Os instrumentos de representação da informação para uso em ambiente digital têm sido agrupados sob a designação de Sistemas de Organização do Conhecimento (SOC). Nesse ambiente, os SOCs “são vistos como esquemas que visam organizar, gerenciar e recuperar informações” (VICKERY, 2007, *on-line*). Dessa forma, eles têm uma única finalidade: “organizar conteúdos para apoiar a recuperação de itens relevantes, disponibilizados na base de dados de uma biblioteca digital” (HODGE, 2000, p. 9). Para cumprir essa finalidade, é importante imprimir maior riqueza semântica na estrutura conceitual dos SOCs.

Tradicionalmente, o tesauro é um tipo de SOC que possui uma estrutura conceitual semântica, pois é composto por um conjunto de conceitos que estão ligados entre si por diferentes tipos de relações. Esses relacionamentos formam uma rede semântica de relações entre termos e conceitos, representados por relações de equivalência (termos sinônimos e variações linguísticas), identificados pelos símbolos USE e UP (usado para), hierárquicas (conceitos ordenados e agrupados por níveis diferentes de generalidade e especificidade), identificados pelos símbolos BT (termo genérico; *broader term*) e NT (termo específico; *narrower term*), e associativas (outras relações que não de equivalência ou hierárquicas), identificadas pelo símbolo RT (termo relacionado; *related term*). Os estudos sobre os princípios aplicados na construção de tesauros vêm evoluindo (MOTTA, 1987; CAMPOS, 1995; CAMPOS; GOMES, 2003, entre outros), e já há iniciativas que propõem a especificação dos diferentes tipos de relações em tesauros.

Entre essas iniciativas, prevalecem as investigações sobre modelos de conversão de tesauros em ontologias de domínio, visando ao reuso do conhecimento já estruturado dos tesauros. Nesse contexto, este artigo relata parte dos resultados da aplicação do modelo desenvolvido por Dagobert Soergel *et al.* (2004) e Lauser *et al.* (2006) na conversão do *Thesaurus* Agrícola Nacional (Thesagro) em um instrumento mais formalizado, de tal forma que os relacionamentos de sua estrutura conceitual estejam explicitados para o usuário. Esse tesauro foi escolhido por ser o único tesauro brasileiro da área da Agropecuária, uma vez que este estudo foi realizado com o apoio da Embrapa Informática Agropecuária (Embrapa), unidade sediada em Campinas-SP, como parte de uma parceria firmada entre essa instituição, a Universidade Federal de Minas Gerais (UFMG) e o Grupo de Pesquisa Protótipo Mapa Hipertextual (MHTX).

2 O modelo de conversão de tesauro

A proposta do modelo utilizado na conversão do tesauro Thesagro tem seus procedimentos detalhados em dois artigos: (1) Soergel *et al.* (2004) e (2) Lauser *et al.* (2006). Esse modelo foi demonstrado pelos autores na conversão de uma amostra do tesauro Agrovoc e teve como objetivo a criação de uma ontologia pesada de domínio, composta por classes, atributos e relacionamentos entre entidades, todos expressos em *Web Ontology Language, Description Logic* (OWL DL). Como a proposta dos autores foi

apresentada de forma gradual, foi possível separar a fase conceitual de modelagem do tesauro da fase de automatização do processo de conversão do tesauro em uma ontologia de domínio. Portanto, os resultados apresentados neste artigo relatam os procedimentos intelectuais de refinamento e explicitação das relações existentes na estrutura conceitual do tesauro Thesagro.

A principal característica desse modelo de reengenharia de tesauros é possibilitar a individualização da modelagem em cinco níveis de entidades: conceito, termo ou lexicalização, *string* ou variantes, notas de escopo e relacionamentos. Dessa maneira, cada nível indica diferentes tipos de informação e os relacionamentos, considerados a espinha dorsal do tesauro, podem ser atribuídos entre entidades de mesmo nível (por exemplo, entre diferentes conceitos) ou entre entidades de níveis distintos (por exemplo, entre termos e *strings*), conforme a seguir:

conceito para conceito	é_uma (hierarquia); praga_de
termo para termo	é_sinônimo_de; é_tradução_de
conceitos para termos	tem_lexicalização (liga os conceitos a seu representante lexical)
termo para *string*	tem_acrônimo; tem_variação_ortográfica; tem_abreviatura (liga os termos com suas formas variantes)

O modelo é composto por três etapas básicas: (1) definição da estrutura do tesauro convertido, utilizando um tesauro existente no domínio a ser trabalhado; (2) coleta de terminologia e outras informações, a partir de um ou mais tesauros, no domínio a ser modelado; (3) edição do tesauro, com a reformulação do tesauro existente, transformando a sua estrutura em uma rede conceitual mais semântica.

2.1 Aplicação das etapas do modelo de conversão de tesauro

Para a aplicação do modelo de conversão de tesauro, foi necessário fazer um planejamento inicial no qual foi constituída uma equipe de trabalho composta por cinco integrantes: um modelador (profissional da informação), um especialista do domínio, dois terminólogos e um tecnólogo.

A primeira etapa, da definição da estrutura do tesauro convertido, abrangeu o mapeamento das características do Thesagro. Verificamos que esse tesauro possui cerca de 9.400 termos descritores, preferidos e não-preferidos, e ele é considerado um

tesauro de alta especificidade e de amplo escopo. De forma semelhante a outros tesauros tradicionais, o Thesagro apresenta os três relacionamentos básicos, utilizando símbolos na língua inglesa:

Relações de equivalência:	USE e USED FOR (UF)
Relações hierárquicas:	BROADER TERM (BT) e NARROWER TERM (NT)
Relações associativas:	RELATED TERM (RT)

Pela análise do tesauro Thesagro, observamos a presença de cerca de 2.000 descritores preferidos não pertencentes a qualquer *cluster* na estrutura hierárquica do Thesagro, sendo considerados descritores órfãos. Essa falta de vinculação entre conceitos pode, algumas vezes, tornar a semântica da estrutura do tesauro menos compreensível para o usuário.

Também na primeira etapa incluiu a determinação da subárea da Intensificação Agropecuária como o recorte temático para a modelagem. A conceitualização desse tema segue da teoria de Boserup (1965) segundo a qual existe uma relação entre a dinâmica da população de uma região, com seu crescimento ou não, o meio ambiente e a utilização de tecnologia na produção agropecuária, originando o aumento ou a manutenção de uma mesma em um dado tempo e lugar. A modelagem dessa subárea teve por base inicial uma taxonomia elaborada pelos especialistas da Embrapa, composta por cerca de 600 conceitos e que foi estruturada em nove categorias: (1) agricultura extensiva; (2) agricultura intensiva; (3) material e métodos; (4) ambiente; (5) agronomia; (6) território e paisagem; (7) socioeconomia; (8) espaço e tempo; (9) instituições. Entre os conceitos da taxonomia foi definida uma amostra composta por 30 conceitos representativos de cada uma das nove categorias e, assim, da subárea da Intensificação Agropecuária como um todo.

Na segunda etapa, da coleta de terminologia, foram utilizados como insumos terminológicos a taxonomia e três tesauros com escopo da agropecuária: o Thesagro, o Agrovoc e o *National Agricultural Library* (NAL). Para essa atividade foi realizada uma comparação entre os conceitos da amostra e a terminologia existente na taxonomia e em cada um dos três tesauros selecionados. Isso foi feito a partir de duas listas: (1) lista composta pelos 30 conceitos da amostra, em português brasileiro, adicionando-se as expressões desses termos no singular e plural; (2) lista composta pelos 30 conceitos da

amostra, traduzidos para o inglês, adicionando-se as expressões dos termos no singular e plural, e também na sua forma inversa (adjetivo + substantivo), por essa inversão ser comum na língua inglesa. A comparação terminológica foi efetivada de forma semiautomática, a partir de uma análise intelectual dos instrumentos e de um processamento automático de comparação. Esse processamento foi realizado utilizando o Extrator de Termos e Estruturas Conceituais Agrícolas Multilíngue (ETECAM), que permitiu resgatar os termos coincidentes com as duas listas nos instrumentos, assim como seus *clusters* semânticos, separadamente.

A terceira etapa envolveu a edição da remodelagem da estrutura conceitual do Thesagro, utilizando o *software* Termos Eletrônicos (e-Termos), que é uma ferramenta que dá suporte à criação e gestão de produtos terminológicos para distintos fins (ensino, glossários, vocabulários controlados). O sistema e-Termos foi desenvolvido como um ambiente computacional colaborativo *web*, que pode ser utilizado de forma gratuita, mas restrita a usuários cadastrados. Ele é composto por seis etapas, com procedimentos automatizados e semiautomatizados, tendo por base os princípios da Teoria Comunicativa da Terminologia (TCT), desenvolvida por Cabré (1999). A TCT possui um construto teórico que defende uma semelhança entre os sistemas da linguagem de especialidade e das línguas gerais, pois ambas são regidas pelas mesmas regras e são caracterizadas pelos mesmos fenômenos de sinonímia e variação linguísticas. Esses princípios minimizam a rigidez normativa advinda da Teoria Geral da Terminologia (TGT), desenvolvida por Wüster (1998), cujos fundamentos são geralmente utilizados na construção de tesauros.

O sistema e-Termos possibilitou a criação de uma Base Definicional que compilou e armazenou excertos definitórios explicativos e/ou sobre os conceitos da amostra. O conteúdo desses excertos auxiliou a confecção de um glossário com as definições dos conceitos da amostra. Além disso, o e-Termos também permitiu a elaboração de fichas terminológicas para esses conceitos, que foi composta por 38 campos semânticos, entre os quais destacamos os campos para as definições dos conceitos (do especialista, modelador e final), para informações enciclopédicas e de glosa, para notas de escopo e para os termos em relação de equivalência, de variação linguística, assim como campos para os conceitos em relação hierárquica e associativa.

Na construção do sistema de conceitos para os conceitos da amostra e seus *clusters* semânticos foram utilizados 44 diferentes tipos de relações. Isso gerou um desdobramento que totalizou a representação de cerca de 600 relacionamentos: de gênero e suas espécies, do todo e suas partes, de equivalências, de strings (variações) e associativas.

Para uma exemplificação da estrutura atual do Thesagro e a propostas de reformulação utilizando o modelo de conversão de tesauros, a seguir apresenta-se a atual modelagem do descritor GATO no Thesagro:

GATO *BT* MAMÍFERO DOMÉSTICO *NT* GATO ANGORÁ *NT* GATO DO MATO *RT* FELIS CATTUS DOMESTICUS *RT* FELIS DOMESTICA	FELIS CATTUS DOMESTICUS *RT* GATO FELIS DOMESTICA *RT* GATO

Na estrutura atual do Thesagro nota-se, por exemplo, que o descritor MAMÍFERO DOMÉSTICO é um conceito mais genérico e que os descritores GATO ANGORÁ e GATO DO MATO são conceitos mais específicos de GATO. No que diz respeito aos outros relacionamentos, é possível perceber que o descritor preferido GATO está em uma relação associativa com os descritores preferidos FELIS CATTUS DOMESTICUS e FELIS DOMESTICA, sem que esteja claro o tipo específico de relação associativa.

Aplicando o modelo de conversão de tesauros, a reformulação da modelagem do descritor GATO ficou com a seguinte configuração:

GATO *temTermoGenérico* MAMÍFERO DOMÉSTICO *temTermoEspecífico* GATO ANGORÁ *temTermoEspecífico* GATO DO MATO *temNomeCientífico* FELIS CATTUS DOMESTICUS *temNomeCientífico* FELIS DOMESTICA	FELIS CATTUS DOMESTICUS *temNomePopular* GATO FELIS DOMESTICA *temNomePopular* GATO

Com essa reformulação, o tipo de relação que foi estabelecida entre os conceitos ficou explícito, facilitando o entendimento da estrutura semântica. Assim, qualquer usuário pode apreender que o descritor MAMÍFERO DOMÉSTICO é um termo mais genérico e está em uma relação de gênero-espécie com GATO que, por sua vez, está em uma relação também de gênero-espécie com GATO ANGORÁ e GATO DO MATO.

Ademais, agora percebemos claramente que os descritores FELIS CATTUS DOMESTICUS e FELIS DOMESTICA são nomenclaturas científicas do descritor GATO, que é um nome popular deste animal. Dessa forma, podemos afirmar que a modelagem atual do Thesagro deixa perdida a função semântica que conecta o descritor GATO com os descritores que representam as nomenclaturas científicas. Esse tipo de estruturação se repete na modelagem de todos os organismos vivos incluídos na terminologia do atual Thesagro e explica a existência dos cerca de 2.000 descritores considerados órfãos. Com essa forma de modelagem, a representação atual do Thesagro mantém disperso o sentido semântico da sua estrutura conceitual, dificultando a compreensão do conhecimento do domínio, a não ser quando o usuário já possui esses conhecimentos *a priori*

Com a reformulação dos conceitos da amostra e de seus *clusters* semânticos foram representadas 286 ocorrências de relações hierárquicas, sendo 225 relações de gênero-espécie (com 52 termos gerais e 173 termos específicos) e 61 relações todo-partes (com 22 termos gerais e 39 termos específicos), o que mostra que a nova estrutura do Thesagro mantém a sua característica de especificidade temática. Além disso, foram representados 232 relacionamentos associativos, confirmando que a subárea da Intensificação Agropecuária tem natureza complexa, com muitos conceitos inter-relacionados.

Os resultados evidenciaram que a expressão explícita das relações entre os pares de entidades (conceitos, termos, *strings* e notas de escopo) refinou a semântica da estrutura do tesauro convertido, dando subsídios para facilitar a interoperabilidade entre diferentes tesauros ou sistemas. Ressaltamos que os resultados completos pela pesquisa podem ser consultados na tese da autora principal deste artigo.

3 Considerações finais

O desenvolvimento deste estudo teve como ponto de partida o enriquecimento semântico da estrutura conceitual de um tesauro tradicional, visando a torná-lo um tipo de SOC, denominação que abrange os instrumentos que podem ser compreendidos pela máquina.

Durante o percurso de pesquisa, ficou evidente que o trabalho de construção ou de reformulação de um tesauro é bastante complexo, o que exige o envolvimento de distintos profissionais, tais como bibliotecários, terminólogos e especialistas. Além disso, verificamos a importância do atendimento às necessidades do usuário, uma vez que a terminologia de um tesauro, ainda que seja uma linguagem de especialidade, deve priorizar a linguagem de busca do usuário.

Percebemos, também, que o desenvolvimento da estrutura semântica de um tesauro deve ser orientado por normas e padrões internacionais para construção de tesauros, pois eles estabelecem princípios para uma representação mais formal, o que pode auxiliar na interoperabilidade entre diferentes vocabulários e sistemas.

Consideramos como principal contribuição desta pesquisa a validação de um modelo de conversão de tesauros tradicionais, que pode ser aplicado como solução de integração de dados em diferentes sistemas. Avaliamos que isso pode impactar positivamente na área da Organização da informação, fazendo avançar os estudos acerca de metodologias que podem ser utilizadas na construção de tesauros.

Referências

BOSERUP, E. **The conditions of agricultural growth:** the economics of agrarian change under population pressure. Chicago: Aldine, 1965.

CABRÉ, M. T. **La terminología:** representación y comunicación. Barcelona: Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra, 1999.

CAMPOS, M. L. A. Linguagens documentárias: núcleo básico de conhecimento para seu estudo. **R. Esc. Biblioteconomia UFMG**, Belo Horizonte, v. 24, n. 1, p. 52-62, jan./jun. 1995.
CAMPOS, M. L. A.; GOMES, H. E. Organização de domínios de conhecimento e os princípios ranganathianos. **Perspect. Ci. Inf.**, Belo Horizonte, v. 8, n. 2, jul./dez. 2003.

DAHLBERG, I. Teoria do conceito. Tradução Astério Tavares Campos. **Ci. Inf.,** Rio de Janeiro, v. 7, n. 2, p. 101-107, 1978.

LAUSER, B. et al. From AGROVOC to the Agricultural Ontology Service: Concept Server an OWL model for creating ontologies in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 2006, Colima, Mexico. **Proceedings…** México: DCMI, 2006.

MOTTA, D. F. da. **Método relacional como nova abordagem para a construção de tesauros.** 1987. 89f. Dissertação (Mestrado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Rio de Janeiro, 1987.

SOERGEL, D. et al. Reengineering thesauri for new applications: the AGROVOC example. **Journal of Digital Information**, v. 4, n. 4, 2004.

WÜSTER, E. **Introducción a la teoría general de la terminología y a la lexicografía terminológica.** Barcelona: IULA, 1998.

Modelo conceitual de inteligência competitiva organizacional

Thiciane Mary Carvalho Teixeira
Universidade Estadual do Ceará
thiciane.teixeira@uece.br

Marta Lígia Pomim Valentim
Universidade Estadual Paulista
valentim@valentim.pro.br

1 Introdução

Estratégias organizacionais como a busca sistemática de informações externas e internas que possibilitam aos gestores conhecerem melhor a própria empresa, o mercado, a concorrência e os clientes, e a análise dos dados visando selecionar, compartilhar e disseminar o conhecimento tangível [explícito] e, a partir de sua apropriação, construir conhecimento intangível [tácito], tornou-se diferenciais em ambientes empresariais competitivos. O conhecimento, a informação e as tecnologias de informação e comunicação (TIC) tornaram-se ativos significativos para gerar inteligência competitiva organizacional (ICO).

A partir da prospecção e da análise de dados e informações é possível filtrar e tratar conteúdos relevantes para o uso da organização, cujo valor agregado a partir da análise é direcionando à tomada de decisão e ao planejamento de ações estratégicas. É possível transformar informação e conhecimento em inteligência, cujo objetivo é proporcionar a melhoria contínua do desempenho organizacional, propiciando solucionar problemas, criar novos produtos, projetar mudanças e inovação.

A ICO se refere a um processo estratégico que enfoca a conversão de dados e informações sobre capacidades, vulnerabilidades e intenções da concorrência, em vantagem competitiva para a organização. É um processo que necessita ser construído, mantido e avaliado constantemente, porquanto é movido por uma dinâmica que envolve a busca, o compartilhamento e o uso de informações de modo criativo, visando descobrir

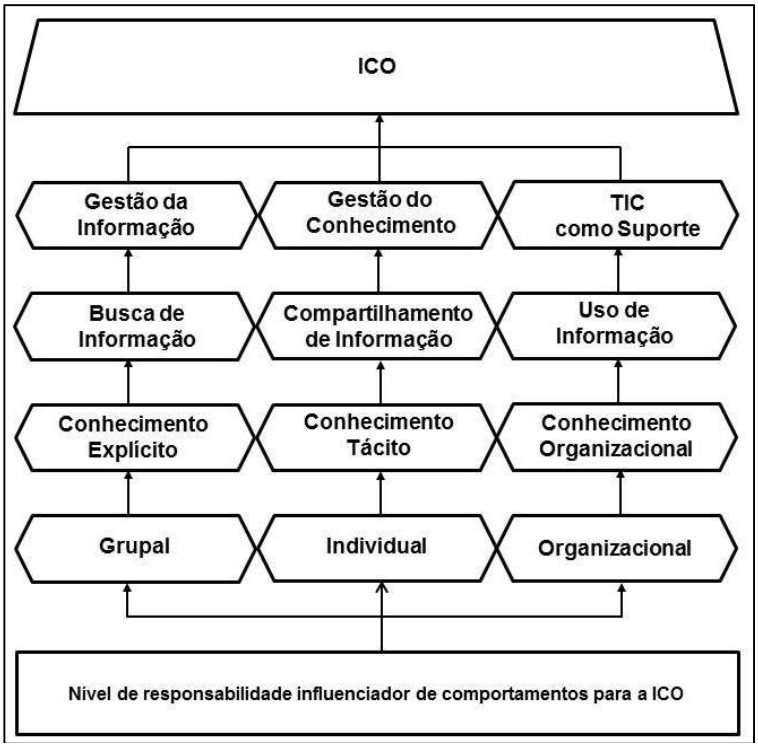
oportunidades, reduzir riscos e propiciar diferentes tipos de inovações combinadas envolvendo qualidade, produtividade e competitividade.

2 Nível de responsabilidade e comportamentos aplicados à ICO

A proposta salienta a identificação de três níveis diferentes de responsabilidades que influenciam os comportamentos de busca, compartilhamento e uso de informação e de TIC para a ICO:

a) ***Individual***: por meio da valorização dos conhecimentos, competências e habilidades individuais explicitados. Refere-se ao conjunto de conhecimentos tácitos dos gestores individualmente, advindos da educação formal, das experiências vivenciadas e do espírito empreendedor. Evidenciou-se uma relação de influência positiva deste nível de responsabilidade com o comportamento de busca, compartilhamento e uso de informação e de TIC.
b) **Grupal**: por meio da socialização de informações é possível criar uma compreensão de grupo sobre uma dada informação ou problema. Refere-se aos conhecimentos de distintos sujeitos organizacionais explicitados e disponibilizados ao grupo de gestores. Evidenciou-se uma relação de influência positiva deste nível de responsabilidade com o comportamento de busca, compartilhamento e uso de informação e de TIC, e com as sutilezas do grupo em se relacionar, organizar e tratar a informação disponível.
c) **Organizacional**: por meio da compreensão sobre o impacto das ações individuais ou grupais, bem como da responsabilidade em relação ao resultado e à competitividade da empresa. Refere-se ao conhecimento organizacional, incluindo a influência da cultura organizacional e informacional nas práticas de busca, compartilhamento e uso de informação e de TIC. Evidenciou-se uma relação de influência positiva deste nível de responsabilidade com o comportamento de busca, compartilhamento e uso de informação e de TIC como suporte ao planejamento e à tomada de decisão.

Figura 1: Nível de responsabilidade influenciador de comportamentos para a ICO



Fonte: Elaborado pelas autoras - 2014.

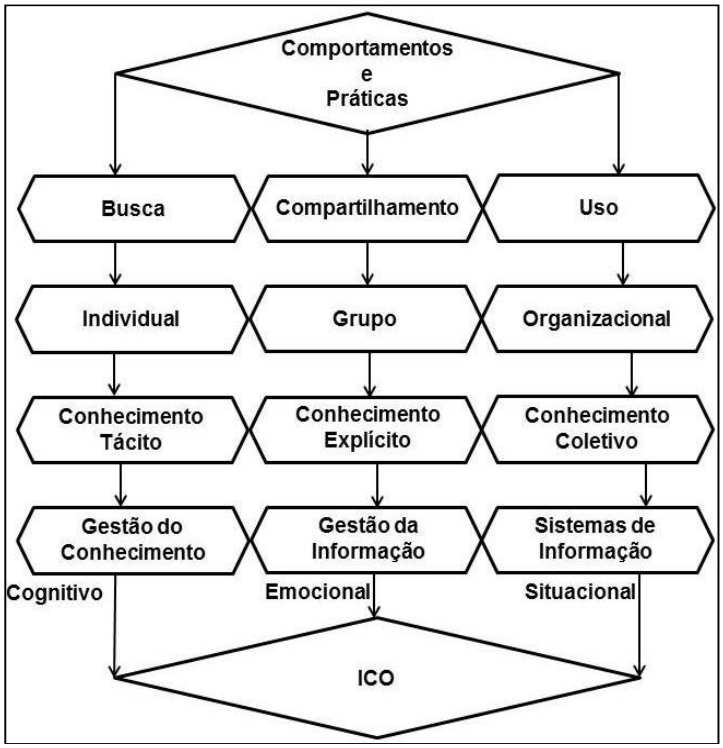
O nível de responsabilidade individual corresponde ao capital humano, refere-se aos conhecimentos tácitos do sujeito organizacional que podem ser externalizados e gerenciados no contexto da organização. O gestor como líder e empreendedor, representa este nível, ou seja, é por meio das interpretações dos fatos e acontecimentos e do compartilhamento de informações é possível gerar novos conhecimentos, que podem ser apropriados por outros sujeitos organizacionais com significados e compreensões próprios à realidade organizacional vivenciada, fator que possibilita a solução de problemas ou a preparação para se antecipar as oportunidades e ameaças externas e internas à empresa.

O nível de responsabilidade do grupo corresponde à construção de significado e compreensões, por meio do entendimento coletivo das informações a ser gerenciado no contexto organizacional. As fontes de informações têm papel imprescindível, uma vez que os fornecedores e as redes associativas participam, geram e fornecem as informações relevantes ao negócio.

O nível de responsabilidade organizacional corresponde à estrutura física e tecnológica que permite o acesso e o uso das informações, e o mais importante a cultura

organizacional e informacional que é imprescindível para garantir a prática de ICO na empresa.

Figura 2: Comportamentos e práticas de ICO



Fonte: Elaborada pelas autoras - 2014

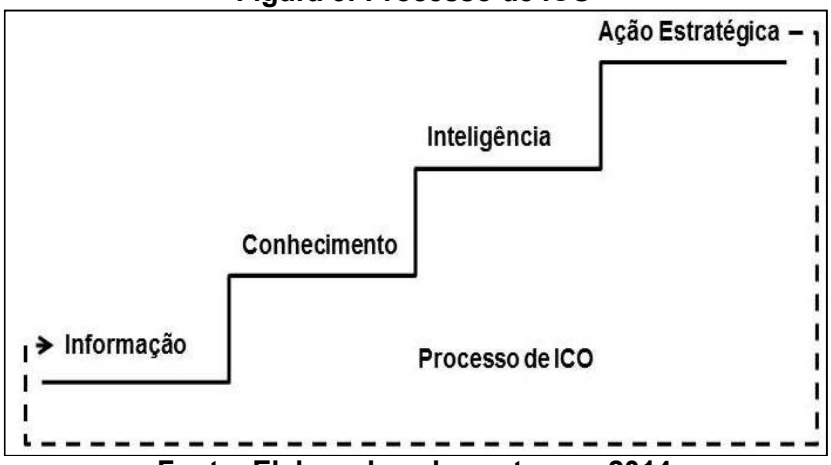
A ICO é um processo de orientação à ação estratégica, seja para a alocação de recursos, para a condução do planejamento de ações futuras, para propiciar inovação em produtos, serviços e processos. A organização precisa desenvolver competências em informação de modo a gerenciar as informações relevantes ao negócio e, também, os conhecimentos gerados pelos sujeitos organizacionais de todos os níveis hierárquicos, visando obter diferenciais competitivos.

3 Processo de ICO

A ICO é compreendida como um processo de aprendizagem, uma vez que as informações precisam ser analisadas e os conhecimentos precisam ser apropriados pelos sujeitos organizacionais, de modo a fazerem inferências em relação à realidade vivenciada pela empresa. A ICO é um processo contínuo que, após a apropriação das

informações relevantes e da internalização dos conhecimentos pelo grupo, se estabelece uma dinâmica de captar, adequar, criar, inovar, perceber e agir de maneira estratégica.

Figura 3: Processo de ICO



Fonte: Elaborada pelas autoras - 2014

A ICO contribui para a obtenção da competitividade por ser um processo contínuo, cuja base reside nos comportamentos e práticas em buscar, compartilhar e usar as informações e as TIC como aprendizagem consistente e sistêmica, alicerçando a crença de que é possível avaliar as perspectivas futuras para inovar, aprender e agir de modo inteligente, mesmo que haja risco em distintos graus quanto à decisão tomada.

O processo de ICO estabelece três orientações que resultam dos comportamentos praticados voltados à ICO (Figura 4):

1) *Orientação para o gerenciamento de informações*

Tem como base as fases 1 e 2 definidas pelo modelo de Miller (2002) que, correspondem a identificação das necessidades e as ações de selecionar e filtrar informações relevantes ao negócio, no ambiente externo e interno. Essa orientação influência positivamente o nível de responsabilidade individual para o comportamento de busca de informação. Destaca-se a importância da motivação e do treinamento para sensibilizar e reconhecer os conhecimentos explícitos presentes nas informações disponíveis no contexto organizacional, com enfoque para o gerenciamento de informações.

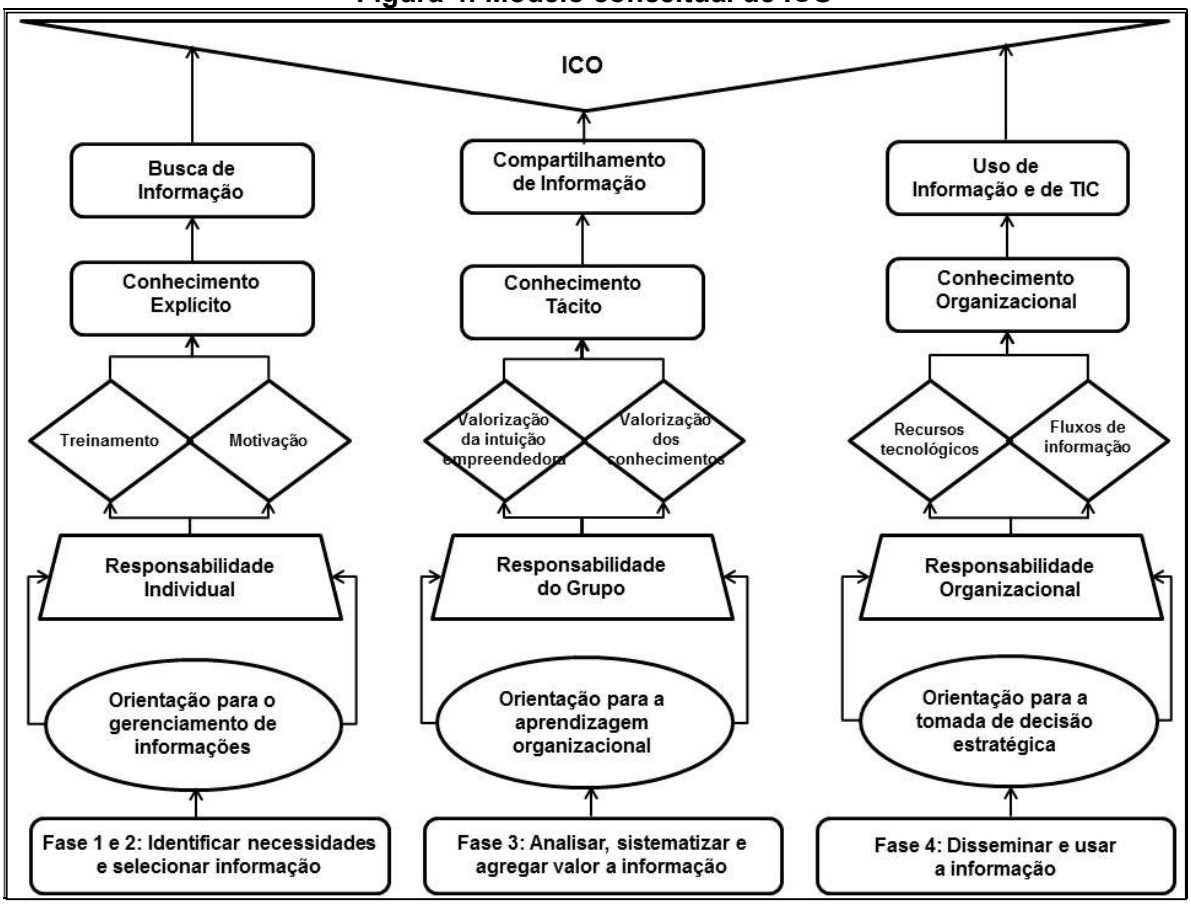
2) *Orientação para a aprendizagem organizacional*

Tem como base a fase 3 definida pelo modelo de Miller (2002) que, corresponde as etapas de análise, sistematização e agregação de valor às informações. Essa orientação influência positivamente o nível de responsabilidade do grupo para o comportamento de compartilhamento de informação. Enfatiza-se a importância da valorização dos conhecimentos tácitos presentes nas intuições e características empreendedoras de cada sujeito organizacional e do grupo para a realização do compartilhamento de informação como alicerce que propiciem a aprendizagem organizacional.

3) *Orientação para a tomada de decisão estratégica*

Tem como base a fase 4 definida pelo modelo de Miller (2002) que, corresponde a etapa de disseminação e uso da informação para gerar inteligência organizacional. Essa orientação influência positivamente o nível de responsabilidade da organização para o comportamento de uso da informação e das TIC para a tomada de decisão estratégica. Ressalta-se a importância da utilização dos recursos tecnológicos como suporte, e dos fluxos de informação e canais de comunicação como estruturas que estimulam a cultura organizacional e informacional para orientar a tomada de decisão.

Figura 4: Modelo conceitual de ICO



Fonte: Elaborada pelas autoras - 2014.

O modelo conceitual proposto de ICO pode ser aplicado a partir das respostas a três questões relacionadas aos comportamentos e práticas de busca, compartilhamento e uso de informação e de TIC:

- *Como a informação chega até a empresa?* Os sujeitos organizacionais necessitam avaliar as fontes de informação quanto à credibilidade para a seleção e filtragem dos conteúdos relevantes. A informação obtida deve ser fidedigna, confiável e consistente, de maneira a servir como insumo ao processo decisório, assim, os gestores precisam acreditar que utilizando as informações disponíveis obterão resultados mais eficazes.
- *Como a informação é compartilhada para a gestão estratégica?* As novas informações apropriadas pelo grupo devem ser adicionadas ao conjunto de

conhecimentos previamente existentes. A intuição e a percepção presentes nos indivíduos, a partir de experiências vivenciadas podem se unir a essas informações, de modo a propiciar condições para avaliar e decidir se podem ser úteis ou não para uma ação estratégica.

- *Como os recursos tecnológicos podem ser aplicados para a tomada de decisão?* O uso de recursos tecnológicos é importante para a busca, o compartilhamento e o uso de informação no contexto organizacional, pois propiciam agilidade e dinâmica ao processo de tomada de decisão.

4 Material e métodos

A pesquisa qualiquantitativa, descritiva e explicativa, aplicou questionário estruturado e roteiro de entrevista, possibilitando a descrição e explicação do fenômeno na tentativa de conhecer e expor as práticas realizadas e o grau de importância dos comportamentos para o processo de ICO.

O método dedutivo partiu das hipóteses: a) os comportamentos de busca e de compartilhamento da informação influenciam positivamente nas atividades de prospecção e monitoramento; b) o uso de informação e de TIC influenciam positivamente nas atividades de tomada de decisão e no planejamento de ações estratégicas.

O universo pesquisado contemplou 29 (vinte e nove) empresas cearenses premiadas em 2012 com o Prêmio Delmiro Gouveia, que estabelece que as maiores empresas devem ser aquelas que possuam um melhor desempenho econômico-financeiro e social no ano, avaliadas de acordo com os faturamentos líquidos: acima de R$ 90 milhões e menor do que R$ 90 milhões.

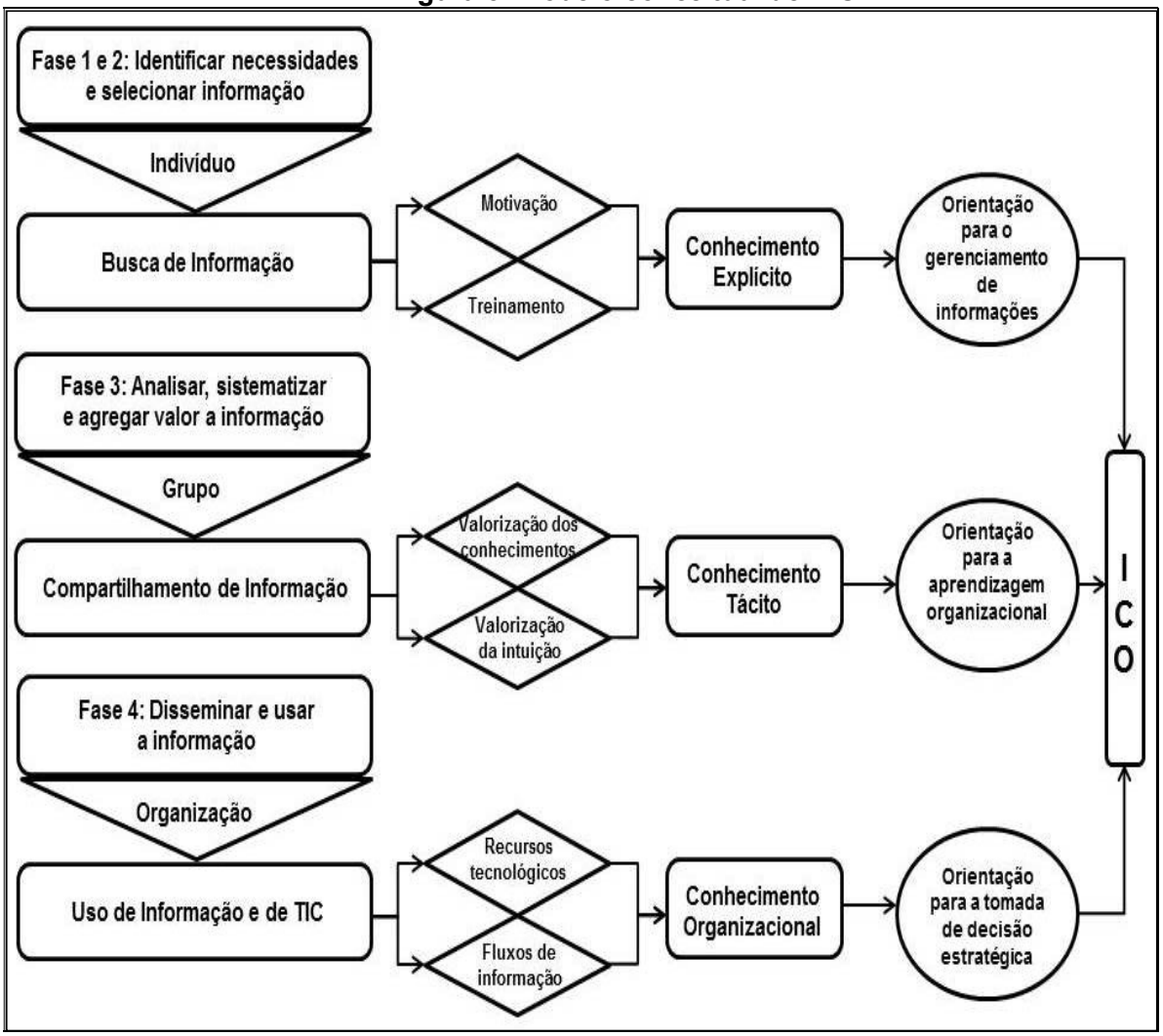
Para a análise de dados, aplicou-se o método ‘Análise de Conteúdo’ de Bardin (1977) e análise estatística simples processada pelo *software* Excell que permitiu apresentar indicadores por meio de gráficos e tabelas.

5 Resultados

O modelo conceitual de ICO (Figura 5) pode servir como diretriz às empresas que atuam em mercados competitivos. Como também pode ser adaptado de maneira realista em empresas de menor tamanho.

A inteligência competitiva organizacional que se deseja alcançar reside no fato de reconhecer que os comportamentos de busca, compartilhamento, uso de informação e de TIC são práticas determinantes para a obtenção de vantagem competitiva.

Figura 5: Modelo conceitual de ICO



Fonte: Elaborada pelas autoras - 2015

A ICO surge como o resultado dos comportamentos e práticas, dos níveis de responsabilidade, e das orientações que podem servir como uma diretriz. Ressalta-se que a cultura organizacional e informacional é fundamental para o processo de ICO, uma vez que os sujeitos organizacionais precisam valorizar a busca, compartilhamento e uso de informação e de TIC para perceber vantagens competitivas.

As hipóteses de que os comportamentos de busca e de compartilhamento da informação influenciam as atividades de prospecção e monitoramento, e que o uso de informação e de TIC influenciam as atividades de tomada de decisão e do planejamento de ações estratégicas foram positivas.

As empresas sabem o que é inteligência competitiva, utilizam e praticam no dia a dia da empresa as etapas de ICO, para a obtenção de diferenciais competitivos. Existem algumas dificuldades para as empresas manterem e sistematizarem as práticas de busca, compartilhamento e uso de informação e de TIC que envolvem variáveis culturais, que emergem como fator importante a ser trabalhado para a aplicação da inteligência competitiva.

As práticas de busca de informação são consideradas estratégias de sobrevivência. Buscar informação junto aos clientes é uma das principais práticas exercidas, visando garantir a qualidade dos serviços e o aumento da competitividade. Evidenciou-se que a busca de informação junto aos clientes é mais constante do que junto aos concorrentes.

A preocupação com as informações internas emerge como uma prática de busca de informação mais sistemática. Há a preocupação com o controle das informações internas, pois influem diretamente o nível tático e operacional.

As atividades de monitoramento se constituem em uma realidade e as formas de busca são variadas. Os monitoramentos político e econômico se mostraram como uma ação essencial. Evidenciou-se que as empresas percebem mais o governo como uma ameaça do que a concorrência de mercado.

A postura empreendedora dos gestores é uma variável importante para a sistematização e agregação de valor às informações, contribuindo para o real e efetivo uso da informação. Após a busca de informação por parte do gestor, o compartilhamento das informações é realizado e influi no coletivo organizacional subsidiando decisões e planejamentos. Além da busca de informações realizada de forma individual no nível estratégico, também se identificou o uso efetivo das TIC nas redes e canais internos organizacionais, por meio dos fluxos de informação propiciando melhorias nos processos, atividades e tarefas diárias.

As empresas fazem uso de sistemas informáticos de informação e comunicação, visando o acesso, a armazenagem e a disseminação de informações para subsidiar o processo decisório e o planejamento, bem como o estabelecimento de ações estratégicas.

6 Considerações

O modelo conceitual de ICO elaborado merece ser validado para ser reconhecido teoricamente. A aplicação do modelo e avaliação das práticas de busca, compartilhamento e uso da informação e de TIC pode ter impacto nos resultados financeiros da empresa.

O estudo propõe ações por meio da motivação, da valorização dos conhecimentos e dos fluxos de informação como estruturas necessárias ao processo de ICO. Os comportamentos de busca, compartilhamento e de uso de informação e de TIC conduzem às práticas empreendedoras, e à aplicação inteligente de informação e conhecimento no formato estratégico, proporcionando impacto nos resultados da organização com vantagem competitiva. A aplicação do modelo de ICO permitirá a otimização das decisões estratégicas, da aprendizagem organizacional e do gerenciamento de informações, com envolvimento e participação do indivíduo, do grupo e da organização.

Referências

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 1977. 226p.

MILLER, J. P. **O milênio da inteligência competitiva**. Porto Alegre: Bookman, 2002. 293p.

TEIXEIRA, T. M. C. **Inteligência competitiva organizacional**: um modelo apoiado nos comportamentos de busca, compartilhamento e uso de informação e de TIC. Marília, 2014. 200f. Tese (Doutorado) – Programa de Pós-Graduação em Ciência da Informação – Universidade Estadual Paulista (Unesp), 2014.

Controle de vocabulário em periódicos científicos eletrônicos

José Carlos Francisco dos Santos
Universidade Estadual de Londrina
jc.fa1982@gmail.com

Brígida Maria Nogueira Cervantes
Universidade Estadual de Londrina
brigidacervantes@gmail.com

1 Introdução

No contexto atual da tecnologia, em se tratando, especificamente, da informação na *web*, os periódicos já nascem no meio eletrônico. Nesse ambiente, o processo de escolha de palavras-chave para a representação de conteúdos temáticos merece cuidado especial, pois são recursos primordiais para a recuperação e acesso aos conteúdos existentes em periódicos científicos eletrônicos.

O estudo tem como objetivo analisar as condições de funcionamento de um controle de vocabulário de palavras-chave atribuídas em periódicos científicos eletrônicos por meio do VCI&I *tools*, como uma ferramenta complementar ao Sistema Eletrônico de Editoração de Revistas (SEER) do IBICT. Justifica-se esta proposição pelos efeitos produzidos para a melhoria da recuperação de conteúdos indexados em periódicos científicos eletrônicos, disponíveis no sistema SEER, além de servir de instrumento de consulta aos autores para a escolha dos termos mais relevantes e consistentes para a disseminação de suas produções científicas.

2 Controle de vocabulário em periódicos científicos eletrônicos

As palavras-chave têm um papel importante para a recuperação dos conteúdos temáticos em publicações científicas. Nas regras de submissão, comumente são vistas

orientações para os autores tomarem como base o texto e executar a atribuição, da melhor forma, como indício para uma representação fidedigna de seu texto (GONÇALVES, 2008).

Na norma internacional ANSI/NISO Z39:19-2005, palavra-chave é definida como sendo a expressão que aparece no texto, configurada por meio da linguagem natural, e é considerada a mais significativa para indexar ou recuperar o texto. (AMERICAN NATIONAL STANDARS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005). Segundo a NBR 6028, palavra-chave é a "Palavra representativa do conteúdo do documento, escolhida, preferentemente, em vocabulário controlado" (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2003b, p. 1). Para Miguéis e Neves (2013), palavras-chave se distanciam do termo de indexação, porque este último tem o resultado centrado no processo de análise conceitual que faz a tradução para o vocabulário controlado e, por seu turno, a atribuição de palavras-chave é livre, sendo escolhida pelo autor, editor ou até mesmo gerada de maneira automática.

Gil Leiva e Alonso Arroyo (2005, tradução nossa) afirmam que o produto dos profissionais de indexação não deve ser substituído pelas palavras-chave atribuídas pelos autores, mas as palavras-chave são uma importante fonte de alimentação para as análises dos artigos científicos nos processos de indexação, assim como uma fonte importante para início das metodologias de uma indexação automática.

O controle de vocabulário geralmente "[...] tem sido exercido por intermédio de: seleção de termos preferidos; distinção de homógrafos; estabelecimento formal das relações entre os termos." (FUJITA, 2011, p. 42). O controle de vocabulário é utilizado para aumentar a eficácia do armazenamento de informações; dos sistemas de recuperação, dos sistemas de navegação na *Web*, e outros ambientes que precisam, tanto para identificar como para localizar o conteúdo desejado de algum tipo de descrição utilizando uma linguagem. O objetivo principal do controle de vocabulário é alcançar consistência na descrição de objetos de conteúdo e facilitar a recuperação (AMERICAN NATIONAL STANDARS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

A necessidade do controle de vocabulário é decorrente de duas características básicas da linguagem natural: 1) duplicidade de nomenclatura para um mesmo conceito – exemplo: SOC/Sistemas de Organização do Conhecimento; 2) duplicidade de conceito para uma mesma palavra – exemplo: manga (camisa), manga (fruta). Para a obtenção do controle de vocabulário, são necessários três principais métodos: a) definir o escopo dos termos; b) usar o relacionamento equivalente para a ligação de termos com sinônimos e quase sinônimos; e c) distinguir entre homógrafo (AMERICAN NATIONAL STANDARS INSTITUTE; NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005).

3 Procedimentos metodológicos

A ferramenta VCI&I *tools* foi desenvolvida na linguagem *PHP*, por se tratar de uma linguagem livre para *web* e realizar uma perfeita integração com o sistema gerenciador de banco de dados *MySQL*. A ferramenta executa os seguintes procedimentos, aqui identificados de requisitos funcionais, enumerados com a sigla RF: RF01 Coleta de metadados; RF02 Análise de Metadados e Vocabulário controlado; RF03 Tratamento Palavras-chave; e RF04 Resultados.

Cabe esclarecer que, para a elaboração da proposta da ferramenta VCI&I *tools*, optou-se pela utilização de vocabulários controlados consolidados e disponíveis eletronicamente, que serão exigidos na funcionalidade de análise de metadados (especificamente as palavras-chave) com vocabulário controlado. Na escolha dos vocabulários controlados para a confirmação de termos, considerou-se, como principal parâmetro, a atualidade e consistência da publicação, sendo eles os oferecidos por Pinheiro e Ferrez (2014) - TBCI e Moreira (2006) - TCI.

O principal objetivo do RF01 é coletar os metadados do periódico selecionado por meio do protocolo OAI-PMH e armazenar na base de dados. O objetivo do RF02 é verificar as palavras-chave que remetem à igualdade dos termos conforme os vocabulários controlados, utilizando os seguintes mecanismos de comparação:

I) idênticas – realiza a comparação entre a palavra-chave e o termo do vocabulário controlado;

II) idênticas retirando as palavras vazias – são excluídas das palavras-chave e dos termos do vocabulário controlado as palavras vazias, por exemplo: artigos definidos e indefinidos (a, o, um, uma), preposições (de, da), entre outras.

III) idênticas retirando os plurais - são tratados os plurais, utilizando o conjunto de regras e exceções de Orengo; Buriol e Coelho (2007), que propõe melhorias para o algorítimo de *stemming* em língua portuguesa RSPL de Orengo e Huyck (2001) - Removedor de Sufixos.

IV) idênticas retirando os sufixos - depende da redução das palavras vazias e dos plurais, as regras e exceções são de Orengo; Buriol e Coelho (2007).

V) idênticas por meio do índice contido - a partir do índice contido a ser calculado e armazenado no banco de dados entre cada palavra-chave e cada termo do vocabulário controlado.

O objetivo do RF03 é realizar o tratamento das palavras-chave não compatibilizadas com os vocabulários controlados com a intervenção do humano para gerenciar este processo. Já o objetivo do RF04 é mostrar os resultados quantitativos dos processamentos referentes à compatibilização de palavras-chave com os vocabulários controlados.

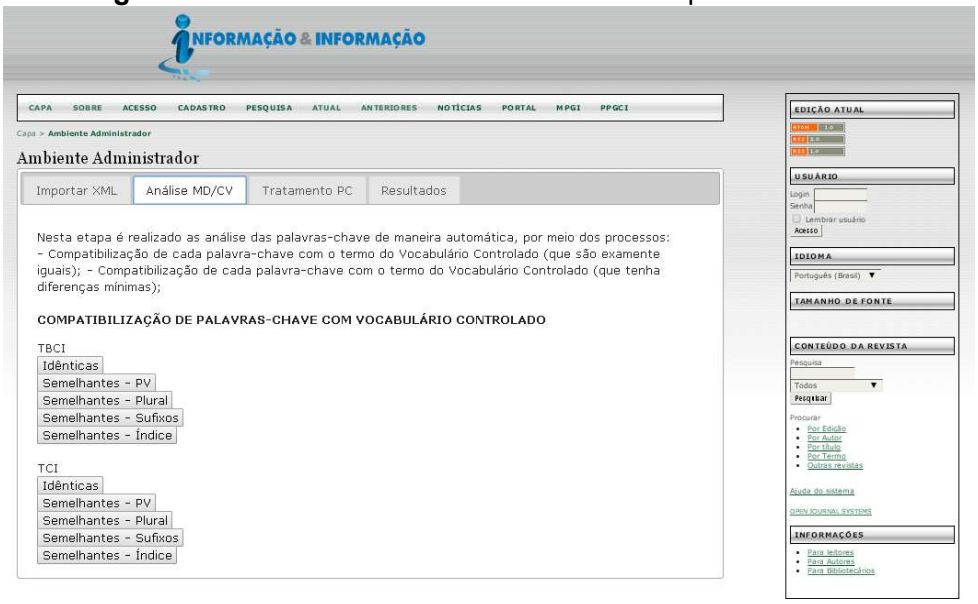
A escolha para a aplicação do protótipo da ferramenta VCI&I *tools* foi a Revista Informação & Informação por ser um periódico científico da área da CI e possuir indicativo de qualidade no Qualis, atualmente B1. A Informação & Informação é vinculada ao Programa de Pós-Graduação em Ciência da Informação, do Departamento de Ciência da Informação da Universidade Estadual de Londrina.

Os dados foram coletados nos metadados dos artigos publicados, desde o primeiro número de 1996 até o último número de 2014. Em uma revista científica, constam vários tipos de seções, como artigos, resenhas de livros/mídia, teses e dissertações, pontos de vista, relatos de experiência, comunicações. Portanto, atribuiu-se a expressão artigos como uma designação genérica para as seções.

4 Aplicação da proposta: análise e resultados

Para alcançar os resultados indicados, executaram-se os processos da ferramenta VCI&I *tools* para a importação dos metadados. Ao todo, a importação consistiu de 437 registros (documentos), com o total de 1199 palavras-chave encontradas, das quais se depurou um total de 365 palavras-chave idênticas, e foram obtidas 834 palavras-chave importadas na base de dados do VCI&I *tools* que serão comparadas no processo de compatibilização. Na Figura 1 visualiza-se a interface de análise das palavras-chave.

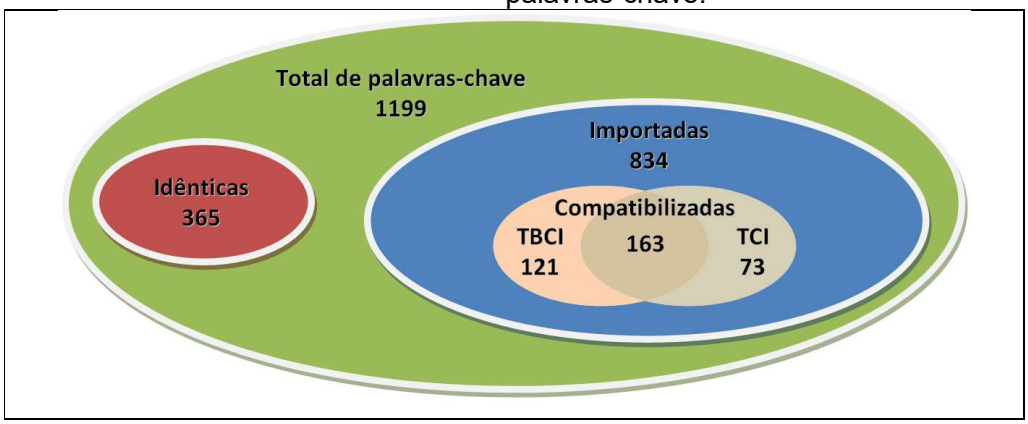
Figura 1 - Interface de análise de metadados - palavras-chave.


INFORMAÇÃO & INFORMAÇÃO
CAPA SOBRE ACESSO CADASTRO PESQUISA ATUAL ANTERIORES NOTÍCIAS PORTAL MPGI PPGCI
Capa > Ambiente Administrador
Ambiente Administrador
Importar XML
Análise MD/CV
Tratamento PC
Resultados
Nesta etapa é realizado as análise das palavras-chave de maneira automática, por meio dos processos: - Compatibilização de cada palavra-chave com o termo do Vocabulário Controlado (que são examente iguais); - Compatibilização de cada palavra-chave com o termo do Vocabulário Controlado (que tenha diferenças mínimas);
COMPATIBILIZAÇÃO DE PALAVRAS-CHAVE COM VOCABULÁRIO CONTROLADO
TBCI
Idênticas
Semelhantes - PV
Semelhantes - Plural
Semelhantes - Sufixos
Semelhantes - Índice
TCI
Idênticas
Semelhantes - PV
Semelhantes - Plural
Semelhantes - Sufixos
Semelhantes - Índice
EDIÇÃO ATUAL
USUÁRIO
IDIOMA
TAMANHO DE FONTE
CONTEÚDO DA REVISTA
INFORMAÇÕES

Fonte: Elaborado pelos autores, utilizando o *layout* do sistema SEER.

O processo de compatibilização realizou-se por meio da interface da ferramenta VCI&I *tools* descritos no RF02. Portanto, como é possível observar, chegou-se a um total de compatibilizações de 121 palavras-chave com somente o TBCI, e de 73 palavras-chave com somente o TCI. O total de palavras compatibilizadas foi 357, apresentados na Figura 2. As palavras-chave não compatibilizadas serão tratadas por processo manual. A partir da compatibilização automática e do tratamento manual das palavras-chave, disponibiliza-se, pelo VCI&I *tools*, o índice de termos.

Figura 2 – Resultados dos processos de importação e de compatibilização das palavras-chave.



Fonte: Elaborado pelos autores.

Nesse índice, apresentado na Figura 3, o usuário tem acesso aos termos preferidos organizados em ordem alfabética, agrupados pela letra inicial dos termos e também por localidades e nomes pessoais. Em cada termo, consta a quantidade de artigos a ele relacionados. Tanto o termo quanto a quantidade de artigos relacionados são *links* em que, ao clicar-se no termo, são apresentadas as palavras-chave a ele relacionadas e as devidas compatibilizações com o TBCI e TCI, se for o caso. Ao clicar na quantidade de artigos relacionados a este termo, o usuário tem, diante de si, o título de cada artigo ligado, por meio de *link*, ao artigo caso queira consultá-lo.

Figura 3 - Interface da lista de termos.

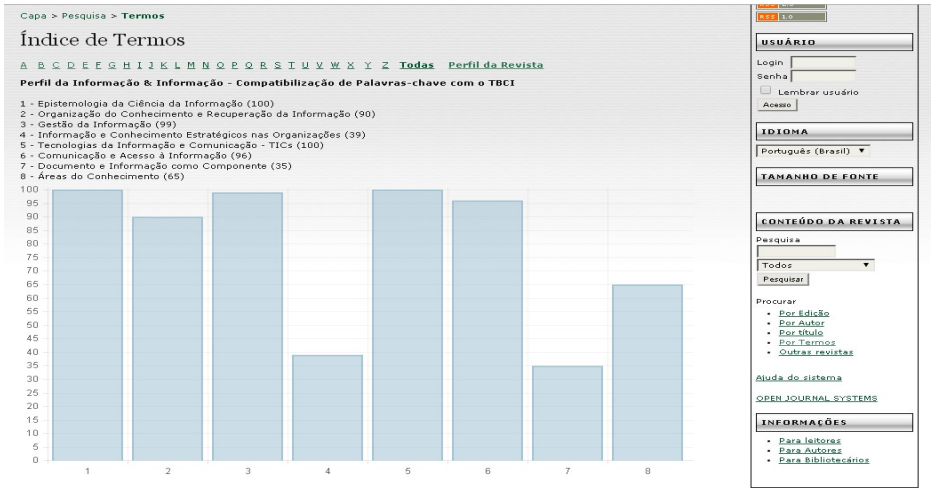
Fonte: Elaborado pelos autores, utilizando o *layout* do sistema SEER.

No ambiente do usuário, é possível acessar o perfil da revista traçado por meio das principais categorias em que se enquadram as palavras-chave do Plano de Classificação Geral do TBCI. O segundo perfil é formado por meio das facetas do TCI. Observa-se que, em ambos os perfis, são apresentados uma lista de categorias/facetas e a quantidade de artigos relacionados por meio da compatibilização de palavras-chave.

Os dados do perfil da revista, de acordo com as categorias do TBCI, são: 1 - Epistemologia da Ciência da Informação (100); 2 - Organização do Conhecimento e Recuperação da Informação (90); 3 - Gestão da Informação (99); 4 - Informação e Conhecimento Estratégicos nas Organizações (39); 5 - Tecnologias da Informação e Comunicação - TICs (100); 6 - Comunicação e Acesso à Informação (96); 7 - Documento e Informação como Componente (35); 8 - Áreas do Conhecimento (65). Este perfil encontra-se representado na Figura 4.

Já os dados do perfil da revista em conformidade com as facetas do vocabulário controlado TCI são: 1 - BI - Informação e operações em biblioteca (73); 2 - CO - Conhecimento e informação (15); 3 - TI - Tecnologia da Informação (64); 4 - AR - Arquivologia (15); 5 - FI - Fontes de Informação (13); 6 - CD - Campos e disciplinas (47); 7 - CI - Ciência da Informação (83); 8 - DC - Documentação (17); 9 - PP - Pessoas, profissionais e grupos formais (38); 10 - PM - Pesquisa e métodos analíticos (19); 11 - MU - Museologia (1); 12 - UI - Unidades de Informação (48); 13 - OR - Organizações (5); 14 - MC - Mídia física e de comunicação (0).

Figura 4 - Perfil da Informação & Informação - Compatibilização de Palavras-chave com o TBCI.


Capa > Pesquisa > Termos
Índice de Termos
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Todas Perfil da Revista
Perfil da Informação & Informação - Compatibilização de Palavras-chave com o TBCI
1 - Epistemologia da Ciência da Informação (100)
2 - Organização do Conhecimento e Recuperação da Informação (90)
3 - Gestão da Informação (99)
4 - Informação e Conhecimento Estratégicos nas Organizações (39)
5 - Tecnologias da Informação e Comunicação - TICs (100)
6 - Comunicação e Acesso à Informação (96)
7 - Documento e Informação como Componente (35)
8 - Áreas do Conhecimento (65)
USUÁRIO
Login
Senha
Lembrar usuário
Acesso
IDIOMA
Português (Brasil)
TAMANHO DE FONTE
CONTEÚDO DA REVISTA
Pesquisa
Todos
Pesquisar
Procurar
Por Edição
Por Autor
Por título
Por Termos
Outras revistas
Ajuda do sistema
OPEN JOURNAL SYSTEMS
INFORMAÇÕES
Para leitores
Para Autores
Para Bibliotecários

Fonte: Elaborado pelos autores utilizando o *layout* do sistema SEER.

Portanto, a revista tem mais publicações nas categorias Epistemologia da Ciência da Informação e Tecnologias da Informação e Comunicação no primeiro perfil; já no segundo perfil, a revista tem mais publicações em CI - Ciência da Informação e BI - Informação e operações em biblioteca.

5 Considerações finais

O controle de vocabulário das palavras-chave de artigos científicos eletrônicos, a compatibilização entre os termos dos vocabulários controlados e as palavras-chave, foram realizados conforme os parâmetros estabelecidos na proposta metodológica, e dentro das expectativas de validação. Para a execução de cada etapa do processo do VCI&I *tools*, tornou-se importante observar a diversidade de sintaxe das palavras-chave atribuídas aos artigos científicos eletrônicos.

O índice de termos proporcionado pela ferramenta, torna-se um meio para conhecer a revista com uma visão holística. O perfil da revista é mais uma opção de visualização do conteúdo por ela tratado e seus vieses de produção científica. Além disso, este modo de exposição da Informação & Informação, têm o objetivo de auxiliar os autores, por meio do índice de termos, na atribuição das palavras-chave às suas produções científicas.

Com o desenvolvimento da ferramenta VCI&I *tools*, haverá a possibilidade de aplicação do procedimento a todos os periódicos da área da Ciência da Informação, gerenciados pelo sistema SEER. Espera-se, também, para futuros estudos, tornar aplicável a ferramenta em outras áreas especializadas gerenciadas por este sistema editoração.

Referências

AMERICAN NATIONAL STANDARS INSTITUTE - ANSI; NATIONAL INFORMATION STANDARDS ORGANIZATION - NISO. **Z39**:19-2005: guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO Press, 2005.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6028**. Informação e documentação: resumo: apresentação. Rio de Janeiro: 2003b.

FUJITA, M. S. L. Modelos de categorização para a construção de tesauros: metodologia de ensino. In: BOCCATO, Vera Regina Casari; GRACIOSO, Luciana de Souza (Org.). **Estudos de linguagem em ciência da informação**. Campinas: Alínea, 2011. Cap. 2.

GIL LEIVA, I.; ALONSO ARROYO, A. La relación entre las palabras clave aportadas por los autores de artículos de revista y su indización en las bases de datos ISOC, IME e ICYT. **Revista Española de Documentación Científica**, Madrid, v. 28, n. 1, 2005. Disponível em: <http://redc.revistas.csic.es/index.php/redc/article/view/165>. Acesso em: 9 set. 2014.

GONÇALVES, A. L. Uso de resumos e palavras-chave em ciências sociais: uma avaliação. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 13, n. 26, p. 1-15, 2008. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2008v13n26p78/6636>. Acesso em: 9 set. 2014.

MIGUÉIS, A.; NEVES, B. Uma abordagem à linguagem de indexação dos artigos científicos depositados no repositório científico da Universidade de Coimbra. **Pontodeacesso**, Salvador, v. 7, n. 1, p. 116-131, abr. 2013.

MOREIRA, M. P. **Tesauro em Ciência da Informação**. Belo Horizonte: PUC Minas. Instituto de Informática; UFMG, [2006]. Disponível em: http://icei.pucminas.br/ci/tci/index.php. Acesso em: 9 mar. 2015.

ORENGO, V. M.; HUYCK, C. (2001). A Stemming Algorithm for the Portuguese Language. 8th International Symposium On String Processing And Information Retrieval (spire), Laguna de San Raphael, Chile, n. , p.183-193, DOI: 10.1109/SPIRE.2001.989755

ORENGO, V. M.; BURIOL, L. S.; COELHO, A. R. (2007). A Study on the Use of Stemming for Monolingual Ad-Hoc Portuguese Information Retrieval. In: CLEF 2006, Alicante, Berlin Heidelberg: Springer-Verlag. v. 4730 p. 91-98. DOI: 10.1007/978-3-540-74999-8_12.

PINHEIRO, L. V. R.; FERREZ, H. D. **Tesauro Brasileiro de Ciência da Informação**. Rio de Janeiro; Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict), 2014. 384 p. Disponível em: http://www.ibict.br/publicacoes-e-institucionais/tesauro-brasileiro-de-ciencia-da-informacao-1/tesauro-brasileiro-de-ciencia-da-informacao/resolveuid/1c60ede36f47aee60c48957ef6db7510. Acesso em: 12 nov. 2014.

As linguagens documentárias na descrição arquivística

Maria de Fátima Santos de Lima
Bibliotecária
fatimasl2006@yahoo.com.br

Francisco Aragão Pedroza da Cunha
Universidade Federal da Bahia
pedrozaici@gmail.com

Introdução

Este artigo é o resultado de Trabalho de Conclusão de curso de Graduação em Arquivologia pela Universidade Federal da Bahia apresentado em 2014.2, que teve como tema a representação da informação nos arquivos a partir da Descrição Arquivista.

A Descrição Arquivística é um processo de pesquisa essencial para o tratamento, representação, recuperação e consequentemente, como ferramenta de disseminação da informação. Logo, é uma atividade primordial para efetivar uma outra atividade no âmbito da Arquivística, a da disseminação da informação.

Desse modo, este trabalho tem como objetivo verificar a utilização de taxonomias e de ontologias para representar a informação orgânica contida nos documentos custodiados pelas instituições arquivísticas públicas estaduais brasileiras.

A natureza da pesquisa foi aplicada com abordagem qualiquantitativa. De acordo com os objetivos, foi classificada como exploratória e descritiva. O universo da pesquisa foram as instituições arquivísticas públicas estaduais brasileiras, e a amostra selecionada foram as instituições que tem sites oficiais e disponibilizam instrumentos de pesquisa para consulta *online*.

Descrição arquivística

Para representar um acervo é necessário o desenvolvimento de atividades e processos que efetivamente permitam o uso e a disseminação da informação. Nesse sentido, destaca-se a atividade de Descrição Arquivística que pode ser considerada como um "processo pelo qual o elaborador da descrição apreende, identifica, condensa e sem distorções, apresenta todas as possibilidades de uso e aplicação da documentação por ele descrita" (BELLOTTO, 2006, p. 178).

Geralmente, a Descrição Arquivística está relacionada apenas à elaboração de instrumentos de pesquisa. Como ressalta Oliveira (2010, p. 43), "a descrição arquivística tem sido apresentada na literatura da área de forma redutora em relação ao que de fato significa e representa, [...] nas últimas décadas, está vinculada ao cumprimento de padrões que visam apenas à elaboração de instrumentos de pesquisa [...]." No entanto, além de elaboração de instrumento de pesquisa, a Descrição Arquivística estabelece o controle intelectual, explica e promove o acesso ao acervo, e cria meios para representar a informação de acordo considerando o contexto de criação dos documentos.

Além de considerar o contexto de criação do documento, a atividade de Descrição Arquivística abrange uma série de processos e atividades, tais como: pesquisa, indexação, organização, elaboração de instrumentos de pesquisa, representação da informação, representação do conhecimento como objetivo de disseminar a informação.

Cabe nesse momento definir representação da informação e representação do conhecimento. Assim, representação da informação, pode ser compreendida como o "conjunto de atributos que representa determinado objeto informacional e que é obtido pelos processos de descrição física e de conteúdo do documento; e a representação do conhecimento se constitui numa estrutura conceitual que representa modelos de mundo e que pode ser concretizada por meio de diferentes tipos de sistemas de organização do conhecimento (SOC)." (BRASCHER; CAFÉ, 2008, p. 6). A minha opção nessa pesquisa é a representação da informação a partir de sistemas de organização do conhecimento.

Para Dodebei (2002) a representação da informação é composta por dois elementos: objetos (i.e. coisas que queremos representar), que no contexto desse

trabalho são os documentos; e propriedades (i.e. características de tais coisas), as palavras ou os descritores pelos quais os documentos serão representados.

A representação das informações por meio de linguagens documentárias subsidia o processo de descrição e a criação de produtos para auxiliar o acesso e a recuperação dessas informações. "Essas linguagens são, pois, construídas para indexação, armazenamento e recuperação da informação e correspondem a sistemas de símbolos destinados a 'traduzir' os conteúdos dos documentos." (CINTRA, *et.al*.2002, p. 16). Sendo assim, facilitam a comunicação, representam as informações de maneira sintética além de proporcionar maior rapidez na busca e recuperação.

Tratamento temático da informação orgânica

O tratamento temático da informação pode ser entendido como o processo que envolve a descrição física e de conteúdo dos objetos informacionais, o produto desse processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico (BRASCHER; CAFÉ, 2008).

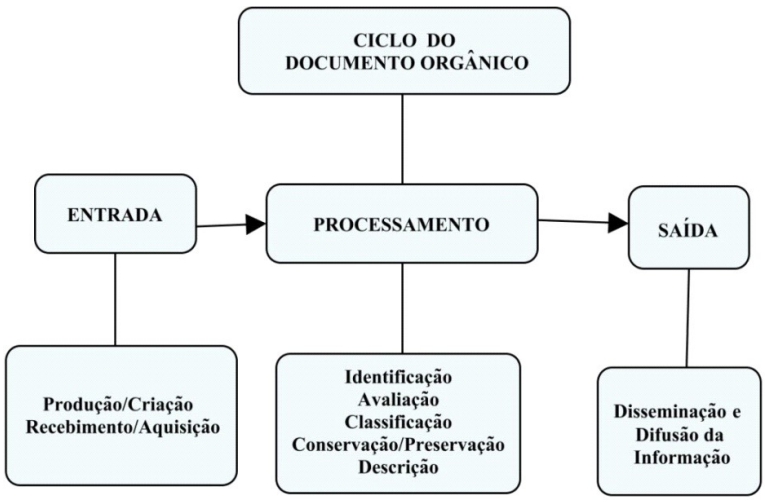
O tratamento temático da informação é responsável pela mediação entre o produtor e o usuário informação, exercendo a tarefa fundamental de dar acesso ao conteúdo documental (NARUKAWA; SALES 2012, p. 154). Esse tratamento é parte de um conjunto de operações denominado de ciclo documentário.

Para Robredo (2005, p. 8), o ciclo documentário "pode ser considerado como um sistema ao qual se incorporam uma série de elementos para serem tratados e convertidos em um produto novo, mais fácil de difundir ou de ser assimilado pelo usuário." Este ciclo envolve três processos básicos: entrada, processamento e saída. Para o mesmo autor, a entrada consiste na seleção e na aquisição do documento, o processamento é o conjunto de atividades relativas ao tratamento técnico do documento, e a saída consiste na busca e recuperação da informação.

No contexto do documento orgânico ou de arquivo, este ciclo está ligado às atividades arquivísticas que para Rousseau e Couture (1998), são denominadas por funções e estão relacionadas ao conjunto de princípios e operações que dizem respeito ao tratamento e a organização dos arquivos. Para esses teóricos, essas funções são as

de criação, aquisição, avaliação, conservação, classificação, descrição e de difusão. Vale salientar que Rodrigues (2011) acrescenta outra função: a de identificação. A Figura 1 ilustra o ciclo do documento orgânico fundamentado em Robredo.

Figura 1: Ciclo do documento orgânico.



Fonte: Elaboração da autora baseada em Robredo (2005).

No ciclo do documento orgânico, a entrada pode ser definida como a etapa de produção/criação em que as organizações, pessoas ou famílias produtoras geram documentos para o mais variados fins. Já o recebimento ou aquisição, se concretiza quando os acervos documentais são recolhidos ou transferidos para as instituições arquivísticas ou para os serviços de arquivos permanentes, categorizando, assim, as organizações custodiadoras de documentos.

Todas as etapas desse processo são importantes. No entanto, apenas a Descrição é o objeto deste estudo. Na atividade de Descrição Arquivística, vale destacar a importância do uso de linguagens documentárias - as ontologias e as taxonomias - durante o processo de indexação, por contribuir para a qualidade dos resultados das operações de acesso e da recuperação dos documentos orgânicos.

Indexação na área de Ciência da Informação pode ser entendida como o processo pelo qual o indexador identifica os conceitos de que trata um determinado documento e os expressa por meio de descritores ou palavras-chave. Para Robredo (2005, p. 165), a indexação consiste em indicar o conteúdo temático de uma unidade de informação, mediante atribuição de um ou mais termos (ou códigos) ao documento de forma unívoca.

Existem duas maneiras de se indexar documentos: a indexação por palavras e a indexação por conceito. A primeira consiste em utilizar os termos do próprio autor, palavras encontradas no título ou no texto dos documentos, ou seja, a linguagem natural. Já a segunda maneira, a indexação por conceito, é realizada a partir das palavras e suas associações. Procede-se identificando os conceitos e, assim, estes são codificados ou expressos em forma de descritores, configurando a linguagem controlada (ROBREDO, 2005).

A indexação por conceitos exige um grau maior de cognição por parte do indexador e mais tempo para a efetivação do processo. Esse tipo de indexação amplia as possibilidades de recuperação da informação, pois padroniza a terminologia e garante maior precisão e eficácia na busca da informação.

Nos arquivos, o processo de indexação faz parte da Descrição Arquivística que, por sua vez, observa as características e contextos dos conjuntos documentais, ou seja, "a unidade da análise documentária será um fundo; série e subséries, respeitando o Princípio da Proveniência." (AGUIAR, 2008, p.184).

O Princípio da Proveniência e da Ordem Original ("Princípio segundo o qual o arquivo deve conservar o arranjo dado pela entidade coletiva, pessoa ou família que o produziu" (ARQUIVO NACIONAL, 2005, p. 137)), direcionam todo o processo de tratamento documental nos arquivos.

No entanto, apesar de todos os esforços da Descrição Arquivística, a busca e recuperação de documentos não são satisfatórias. A falta de precisão e a alta revocação dos descritores atribuídos aos documentos durante o processo de indexação, e o aumento da oferta de acervos através de portais e sites na web requer um padrão de vocabulário. Para Souza e Alvarenga (2004, p. 133), a web foi implementada de maneira quase anárquica e caótica. Esses autores afirmam que na web:

> [...] não há nenhuma estratégia abrangente e satisfatória para a indexação dos documentos nela contidos, e a recuperação das informações, possível por meio dos "motores de busca", é baseada primariamente em palavras-chave contidas no texto dos documentos originais, o que é muito pouco eficaz, (SOUZA; ALVARENGA, 2004, p. 133).

A polissemia e a ambiguidade dos termos usados na indexação dos documentos são outros fatores que dificultam a recuperação da informação na web devido a deficiência na utilização de padrões semânticos. Assim surge a web semântica:

> [...] a web semântica é uma extensão da web atual, porém apresenta estrutura que possibilita a compreensão e o gerenciamento dos conteúdos armazenados na web independente da forma em que estes se apresentem, seja texto, som, imagem e gráficos à partir da valoração semântica desses conteúdos, e através de agentes que serão programas coletores de conteúdo advindos de fontes diversas capazes de processar as informações e permutar resultados com outros programas (BERNERS-LEE, 2001, *apud* RAMALHO, 2007, p. 3).

Ao acrescentar semântica a um conteúdo, a web semântica irá determinar o sentido de um termo no contexto de um determinado documento. Sendo assim, facilita a construção de aplicações visando diminuir dificuldades relacionadas à recuperação da informação e cria um ambiente de cooperação (PICKLER, 2007).

Esse ambiente da cooperação pode ser possível com a adoção de conceitos padronizados e controle das terminologias usadas em diversas áreas do conhecimento; "visando facilitar a recuperação da informação, diversas pesquisas estão sendo feitas ao longo dos últimos anos em torno dos termos taxonomia e ontologia. Especialistas de diversos campos de conhecimento propõem formas dinâmicas de organização da informação [...]." (NOVO, 2010, p. 132).

Ontologias são usadas também em Inteligência Artificial, Ciência da Computação, Web Semântica e Ciência da Informação como representação do conhecimento e da informação. "Na representação da informação, pode entender-se como a soma de uma série de conceitos relevantes que representam o conhecimento compartilhado pelos membros de um determinado domínio." (MOREIRO GONZÁLEZ, 2011, p. 76).

O uso de ontologias facilita o compartilhamento de informações, define conceitos básicos e o relacionamento entre eles, define os termos para descrever e representar determinada área do conhecimento e atribui sentido e significado ao conteúdo dos documentos.

> Ontologia é um conjunto de conceitos padronizados onde termos e definições devem ser aceitos por uma comunidade no âmbito de um domínio e tem por finalidade permitir que múltiplos agentes compartilhem conhecimento. (CAMPOS, 2007, p. 2).

Um dos principais objetivos das ontologias é o intercâmbio do conhecimento e sua representação. Para isso, os conceitos precisam estar contextualizados e relacionados. Para Rodríguez Barquín e outros(2006), as ontologias apresentam algumas características que têm relação entre si como mostra a Figura 2:

Figura 2: Relações entre classes, propriedades e instâncias.

Fonte: Rodríguez Barquín*et. al*. (2006, p. 6).

Se infere que as classes são representadas pelos conceitos de acordo com o contexto analisado. As instancias são os valores atribuídos a um conceito em uma situação específica. E as propriedades são as características dos conceitos. Nesse sentido, as ontologias são compostas por conceitos estruturados de forma hierárquica e possui características semelhantes às taxonomias.

Do mesmo modo que as ontologias, as taxonomias também são usadas para representar a informação. No campo da Ciência da Informação, a taxonomia está sempre relacionada à organização e a representação da informação.

Para Currás (2010, p. 68), taxonomias são "[...] estruturas conceituais adequadas para seu uso na web semântica e ainda como [o termo] que cria ordem no caos de unidades conceituais, advindas de documentos extremamente variados".

A taxonomia é uma ferramenta que permite classificar e facilitar o acesso a informação em um ambiente desorganizado e desordenado como a internet. Para Moreiro González (2011), no campo das linguagens documentárias, trata-se de uma lista de termos preferenciais com estrutura hierárquica (árvore), que utiliza a relação pai-filho.

Uma estrutura hierárquica reúne termos de determinada área do conhecimento que tem relações entre si e cada termo é superior ao termo seguinte. O fundo está relacionado com a seção, que por sua vez, está relacionado a série e todos são provenientes do mesmo fundo.

Tanto as ontologias quanto as taxonomias definem conceitos em domínios de conhecimento para auxiliar a representação da informação contida nos documentos, a fim de promover a recuperação e acesso à informação. Entretanto existem algumas diferenças entre elas, conforme elencadas no Quadro 2:

Quadro 2: Diferenças básicas entre ontologias e taxonomias

CATEGORIAS	ONTOLOGIAS	TAXONOMIAS
Relações entre os conceitos	Aplica-se basicamente a relações hierárquicas.	As relações acontecem em rede conceitual.
Finalidades	Buscam o desenvolvimento de categorias para representar o conhecimento e auxiliam a recuperação da informação.	Buscam um consenso linguístico em áreas específicas para representar o conhecimento e auxiliar a recuperação da informação.
Aplicação	São mais indicadas para organizar e representar informações em ambientes específicos: portais e intranets.	São mais indicadas para áreas do conhecimento que desejam trabalhar temas com especificidade e aprofundamento.
Tipo de relacionamento entre os termos	“Generalização” ou “tipo de”	“Parte de”, “causa efeito”, “localização”, etc.

Fonte: Elaboração da autora baseado em Vital e Café (2011).

As relações em rede, característica das ontologias dão mais qualidade à estrutura conceitual; do ponto de vista semântico, elas são mais aprimoradas e necessitam de um nível intelectual mais complexo com o apoio de especialistas das respectivas áreas de conhecimento para que os conceitos sejam compartilhados e entendidos por todos. E em geral, ontologias são consideradas mais complexas que as taxonomias, pois representam um determinado domínio do conhecimento (VITAL; CAFÉ, 2011).

Nessa perspectiva, vale ressaltar que o uso de ontologias e taxonomias nos sistemas de indexação é fundamental para a busca, a representação e a recuperação da informação orgânica. Porém, esses sistemas de organização do conhecimento ainda são pouco utilizados na área de Arquivologia.

Considerações finais

São muitos os aspectos a serem considerados para uma eficiente busca e recuperação da informação dos documentos orgânicos. Nesse contexto, a Descrição Arquivística se consolida como um processo de pesquisa essencial para o tratamento, representação, recuperação e consequentemente, como ferramenta de disseminação da informação, além de possibilitar a troca de informações, garante ao Arquivo seu papel de difusor de conhecimento.

Assim a Descrição Arquivística não deve ser vista apenas como um conjunto de procedimentos para elaboração de instrumentos de pesquisas, e sim, para potencializar o uso recorrente das informações nos arquivos, e a utilização de ferramentas de organização e representação da informação durante o processo de indexação, como ontologias e taxonomias.

Os resultados indicaram que há uma necessidade latente de mais estudos a respeito das ontologias e taxonomias como ferramentas para organizar e representar o conhecimento e a informação em SOC e principalmente das aplicações nos sistemas de indexação na área de Arquivologia.

Percebemos também a possibilidade de aplicação de ontologias e de taxonomias, como ferramentas fundamentais para organizar e representar a informação em SOC na área de Arquivologia, e que proporcionam ao usuário final a recuperação da informação de forma rápida e segura.

Referências

AGUIAR, F. L. **O controle de vocabulário controlado como dispositivo metodológicopara organização, tratamento e recuperação da informação arquivística.** 267 f. 2008.Dissertação (Mestrado em Ciência da Informação) – Pontifícia Universidade Católica deCampinas - PUCC, Campinas, 2008. Disponível em: <http://www.bibliotecadigital.puc-campinas.edu.br/tde_busca/arquivo.php?codArquivo=437>. Acesso em: 10 set. 2014.

ARQUIVO NACIONAL. **Dicionário brasileiro de terminologia arquivística**. Rio de Janeiro: Arquivo Nacional, 2005. Disponível em: http://www.arquivonacional.gov.br/Media/Dicion%20Term%20Arquiv.pdf. Acesso em: 10maio. 2014.

BELLOTTO, Heloísa Liberalli. O sentido de descrição documental. In:_____. **Arquivos permanentes**: tratamento documental. 4. ed. Rio de Janeiro: FGV, 2006, p. 173-178.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: Encontro Nacional de Pesquisa em Ciência da Informação, 9., 2008. **Anais...** São Paulo: ANCIB, 2008. Disponível em: <http://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHER%20CAF%C3%89(2008)-1835.pdf>. Acesso em: 28 out. 2014.

CAMPOS, Maria Luiza de Almeida. **A problemática da compatibilização terminológica e a integração de ontologias**: o papel das definições conceituais. Grupo Temático Ancib: GT 2: Organização do Conhecimento e Representação da Informação. Disponível em: <http://www.uff.br/ppgci/editais/marialuizaleituras.pdf>. Acesso em: 29 set. 2014.

CINTRA, Ana Maria. et. al. **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002.

CURRÁS, Emília. **Ontologias, taxonomia e tesauros**: em teoria dos sistemas e sistemática. Brasília: Thesaurus, 2010.

DODEBEI, Vera Lúcia Doyle. **Tesauro**: linguagem de representação da memória documentária. Rio de Janeiro: Interciência: Niterói: Intertexto, 2002.

MOREIRO GONZÁLEZ, José Antonio. **Linguagens documentárias e vocabulários semânticos para a web:** elementos conceituais. Salvador: EDUFBA, 2011.

NARUKAWA, Cristina Miyuki; SALES, Rodrigo. A relevância e influência de linguagens documentárias na política de tratamento da informação. In: Gil Leiva, Isidoro; Fujita, Mariângela Spotti Lopes (Org.). **Política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. Disponível em: http://www.marilia.unesp.br/Home/Publicacoes/politica-de-indexacao_ebook.pdf. Acesso em: 10 set. 2014.

NOVO, Hildenise Ferreira. A taxonomia enquanto estrutura classificatória: uma aplicação em domínio de conhecimento interdisciplinar. **Ponto de Acesso**, Salvador, v. 4, n. 2, p. 131-156, set. 2010. Disponível em: <http://www.portalseer.ufba.br/index.php/revistaici/article/viewArticle/4103>. Acesso em: 10 out. 2014.

OLIVEIRA, Lúcia Maria Velloso de. **Modelagem e status científico da descrição arquivística no campo dos arquivos pessoais.** 2010, 188f. Tese (Doutorado em História Social). Universidade de São Paulo – Faculdade de Filosofia, Letras e História Social – Departamento de História. São Paulo, 2010. Disponível em: <http://www.teses.usp.br/teses/disponiveis/8/8138/tde-14062011-134720/pt-br.php>. Acesso em: 15 set. 2014.

PICKLER, Maria Elisa Valentim. Web semântica: ontologias como ferramentas de representação do conhecimento. **Perspectivas em Ciência da Informação**. São Paulo, v. 12, n. 1, p. 65-83, jan./abr. 2007. Disponível em: <www.brapci.ufpr.br/download.php?dd0=15365.> Acesso em: 10 out. 2014.

RAMALHO, Rogério Aparecido Sá. et. al. Web semântica: uma investigação sob o olhar da Ciência da Informação.**Data Grama Zero** - Revista de Ciência da Informação, v.8, n.6 dez/07. Disponível em: <http://www.dgz.org.br/dez07/Art_04.htm>. Acesso em: 10 out. 2014.

ROBREDO, Jaime. **Documentação de hoje e de amanhã**: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas 4ª ed. Brasília: Edição do autor. 2005.

RODRIGUES, Ana Célia. Identificação: uma nova função arquivística?.**Revista EDICIC**, v.1, n.4, p.109-129, Out./Dez. 2011. Disponível em: <http://www.edicic.org/revista/>. Acesso em: 10 set. 2014

RODRÍGUEZ BARQUÍN, Beatriz Ainhize; MOREIRO GONZÁLEZ, José Antonio; PINTO, Adilson Luiz. Construção de uma ontologia para sistemas de informação empresarial para a área de Telecomunicações.**Data Grama Zero: r**evista de Ciência da Informação, v.7 n.2, abr. 2006. Disponível em: < http://www.dgz.org.br/abr06/Art_04.htm> Acesso em: 30 set. 2014.

ROUSSEAU, Jean-Yves; COUTURE, Carol. O lugar da arquivística na gestão da informação. In:_____. **Os fundamentos da disciplina Arquivística**. Lisboa: Dom Quixote, 1998.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A Web Semântica e suas contribuições para aCiência da Informação. **Ci. Inf**., Brasília, v. 33, n. 1, p. 132-141, jan./abril 2004. Disponível em: <http://www.scielo.br/pdf/ci/v33n1/v33n1a16.pdf> . Acesso em: 01 out. 2014.

VITAL, Luciane Paula; CAFÉ, Ligia Maria Arruda. Ontologias e taxonomias: diferenças. **Perspect. ciênc. inf.** Belo Horizonte, v.16 n. 2, abr./jun. 2011. Disponível em: <http://www.scielo.br/scielo.php?pid=S1413-99362011000200008&script=sci_arttext > Acesso em: 14 set. 2014.

As pesquisas em sistemas de organização do conhecimento nas conferências DCMI

Felipe Augusto Arakaki
Universidade Estadual Paulista
fe.arakaki@gmail.com

Plácida Leopoldina Ventura Amorim da Costa Santos
Universidade Estadual Paulista
placida@marilia.unesp.br

Rachel Cristina Vesú Alves
Universidade Estadual Paulista
rachel@marilia.unesp

Introdução

Com as mudanças de suportes proporcionadas pelas Tecnologias, a Organização do Conhecimento sempre esteve disposta a discutir novas possibilidades e melhorias para organização e representar recursos no ambiente da Web.

Nesse sentido, muito de seus pesquisadores contribuíram com o desenvolvimento do *Dublin Core*, um padrão de metadados criado em 1995 para localizar recursos no ambiente da *Web*. Desde a criação do *Dublin Core*, pesquisadores de diversas áreas reúnem em eventos promovidos pela *Dublin Core Metadata Iniciative* (DMCI), instituição responsável por administrar o *Dublin Core*, para discutir questões relacionadas aos metadados no ambiente da *Web*.

Assim, com a proposta de internacionalização do padrão *Dublin Core*, a partir de 2001, a atual estrutura passou a receber apresentação de trabalhos e muitas foram as contribuições de pesquisa referentes aos Sistemas de Organização do Conhecimento (SOC).

> Sistemas de Organização do Conhecimento (SOC) [termo traduzido do inglês para *Knowledge Organization System* (KOS)] os quais abrangem todos os tipos de esquemas que organizam e representam o conhecimento, por exemplo, as classificações, taxonomias, tesauros e ontologias. SOC são sistemas conceituais semanticamente estruturados que contemplam termos, definições, relacionamentos e propriedades dos conceitos. (CARLAN; MEDEIROS, 2011, p. 54).

Nesse período, de 2001 até 2014, foram realizados 14 eventos ao todo, sendo cinco (5) na Europa e cinco (5) Ásia, três (3) na América do Norte e um na América Central.

Nesse trabalho questiona-se qual o panorama das pesquisas relacionadas aos SOC apresentadas nos eventos do DCMI? Dessa maneira, o objetivo é apresentar o estado da arte dos estudos apresentados nos eventos organizados pelo DCMI que abordam de alguma forma o tema de SOC.

2 Metodologia

É uma pesquisa teórica, de caráter qualitativa e bibliográfica, com o intuito de identificar o estado da arte das pesquisas apresentadas nos eventos do DCMI de 2001 até 2014 no idioma em inglês.

Foram selecionados apenas os trabalhos que abordam estudos temáticos de SOC como tesauros, ontologias, vocabulários controlados, entre outros. Os passos foram: I. Identificação dos trabalhos completos apresentados no DCMI (380 trabalhos); II. Seleção dos trabalhos a partir do título, palavras chave e resumo (35 trabalhos); III. Leitura exploratória dos textos; IV. Extração da temática; V. Classificação em categorias das pesquisas; VI. Esquematização do estudo e análise dos resultados.

Nos anos de 2001 até 2004, 2010 e 2014, não foram identificadas pesquisas que abordaram o tema em questão.

3 As pesquisas em Sistemas de Organização do Conhecimento nas conferências DCMI

Entre as contribuições e influências da Organização do Conhecimento no DCMI pode-se citar, por exemplo, o evento ocorrido em Madrid no ano de 2005, que abordou tema principal "*Vocabularies in Practice*". Durante o evento na Espanha, foram expostas diversas perspectivas de implementação de vocabulários controlados além de sugestões e esforços para padronizá-los.

Miles et al. (2005) discutiram uma definição terminológica para o vocabulário controlado. Lee e Sugimoto (2005) tentaram padronizar vocabulários controlados em três domínios. Greenberg et al. (2005) debateram alguns princípios sobre o vocabulário controlado. Pérez Agüera e Sánchez Jiménez (2005) exploraram a iniciativa de inserir um tesauro no modelo *Resource Description Framework* (RDF). Francesconi e Peruginelli (2005) definiram uma padronização dos metadados e de um vocabulário controlado para os artigos jurídicos da Itália. Kurth, Nehler, Silterra (2005) exploraram o caso particular de uma biblioteca ao administrar um vocabulário controlado.

No ano seguinte, em 2006, Colima no Mexico abordou como tema "*Metadata for Knowledge and Learning*". Apresentou questões voltadas principalmente a padronização e o controle de vocabulários. Ángel Marzal, Ruiz Colmenero e Cuevas Cerveró (2006) propuseram um vocabulário controlado para descrição de objetos educacionais. As questões de compartilhamento de vocabulário ambiental controlado são tratadas por Menger e Rüther (2006) que relataram um serviço de rede semântica, desenvolvido pelo *Umweltbundesamt* (UBA), da Agência Federal do Meio Ambiente da Alemanha e está integrado por vários sistemas de informação ambiental e geográfica. Currier et al. (2006) discutiram a necessidade de desenvolvimento e uso de vocabulários que descrevem objetos educacionais no contexto do Reino Unido. Bartolo et al. (2006) abordaram o desenvolvimento do vocabulário controlado da *National Science Digital Library* (NSDL) para facilitar a alimentação do catálogo a partir da captura de metadados *Dublin Core* durante a execução de códigos computacionais. Miles (2006) propôs uma revisão do projeto de *Simple Knowledge Organization System* (SKOS) antes de uma iniciativa de padronização e apresentou três questões relativas ao desenvolvimento do SKOS para

recomendação do *World Wide Web Consortium* (W3C). Harper (2006) abordou as folhas de estilo *EXtensible Stylesheet Language Transformations* (XSLT) para converter registros da *Library of Congress Headings for Subject* LCSH *Authorities* baseados em registros de *Machine Readable Cataloging eXtensible Markup Language* (MARCXML) ou em formatos *Metadata* Authority Description Schema (MADS) em *eXtensible Markup Language* (XML) para documentos RDF de acordo com o projeto SKOS. Nessa mesma concepção, Liang et al. (2006) abordaram a conversão de dados da AGROVOC para um novo sistema baseado na *Web Ontology Language* (OWL). Pattuelli e Norberg (2006) defenderam uma *Ontologydriven* que representariam as relações sociais, familiares, espacial e temporal entre objetos de aprendizagem para facilitar a construção de significados históricos.

Com o tema "*Application Profiles and their Application in Practice*", Singapore sediou a conferência em 2007. Kakali et al. (2007) propuseram uma ontologia para o perfil de aplicação na museologia. Miyazawa (2007) questionou o atual modelo de idiomas no compartilhamento de informações e da necessidade da padronização da representação linguística. Campbell (2007) explorou o processo de identificação das coisas e como e se acoplam a partir de características semelhantes.

No ano seguinte, com trabalhos abordando questões como perfil de aplicação e a *folksonomia*, marcaram presença em Berlin com tema central "*Metadata for Semantic and Social Applications*" no ano de 2008. Catarino em parceria com Baptista (2008), propuseram um perfil de aplicação social baseado na *folksonomia*. Com uma temática semelhante, Chen, Liu e Qin (2008) abordaram uma metodologia para gerar metadados baseados na extração de *tags*. Kim et al. (2008) mapearam o estado da arte da ontologia das *tags*. Ainda, Lauser et al. (2008) discutiram a criação e elaboração de tesauros por humanos e não humanos (automaticamente). Summers et al. (2008) abordaram uma ferramenta que transforma um cabeçalho de assunto da LCSH em MARCXML para o vocabulário SKOS em RDF.

A cidade de Seul, na Coréia do Sul, sediou em 2009 o tema "*Semantic Interoperability of Linked Data*". Mason (2009) abordou a gestão do conhecimento e o *Dublin Core* e as relações estruturais de modelos conceituais de sistemas de organização do conhecimento que influenciam a eficácia e o fluxo de conhecimento. Panzer e Zeng

(2009) propuseram uma recomendação de uma modelagem de classificação no sistema SKOS. Jeong (2009) defendeu o uso das *tags* para auxiliar na representação de vídeos no YouTube.

Posteriormente, o evento ocorreu na cidade de Hague, na Holanda, e abordou o tema "*Metadata Harmonization: Bridging Languages of Description*", no ano de 2011. No campo temático, Morshed et al. (2011) trabalharam com tesauros no *Linked Data*. Freire, Borbinha e Caldo (2011) buscaram o alinhamento dos SKOS com uma ontologia geográfica. Montiel-Ponsoda et al. (2011) abordaram uma perspectiva de uma padronização para nomenclatura e rotulagem de ontologias para diversas línguas na *Web*. Zavalina (2011) fez uma comparação entre textos livres e nível de coleção dos assuntos em larga escala de bibliotecas digitais.

No ano de 2012, de Kuching, Sarawak, Malaysia com o tema "*Metadata for Meeting Global Challenges"*. Entre os trabalhos apresentados, Ambiah e Lukose (2012) propuseram uma ferramenta para gerar automaticamente páginas com informação semântica a partir de palavras-chave do documento. Zumer, Zeng e Hlava (2012) apresentaram um modelo de domínio para descrever e acessar recursos SKOS. As ontologias foram resgatadas por Bosch et al. (2012) que abordaram o desenvolvimento de uma iniciativa de documentação de dados, definindo, assim, um padrão internacional para documentação e gestão de dados na área de Ciências Sociais, Comportamentais e Econômicas.

A cidade de Lisboa em Portugal apresentou o tema "*Linking to the Future*" em 2013. Os trabalhos de Tian et al. (2013) discutiram um perfil de aplicação e de ontologia baseados no *Linked Data* para música tradicional chinesa. Hillmann, Dunsire e Phipps (2013) verificaram as mudanças de terminologia de diversos padrões de metadados. Nessa mesma vertente, Honma, Nagamori e Sugimoto (2013) abordaram a localização e combinação de vocabulários para serem projetados em perfis de aplicação.

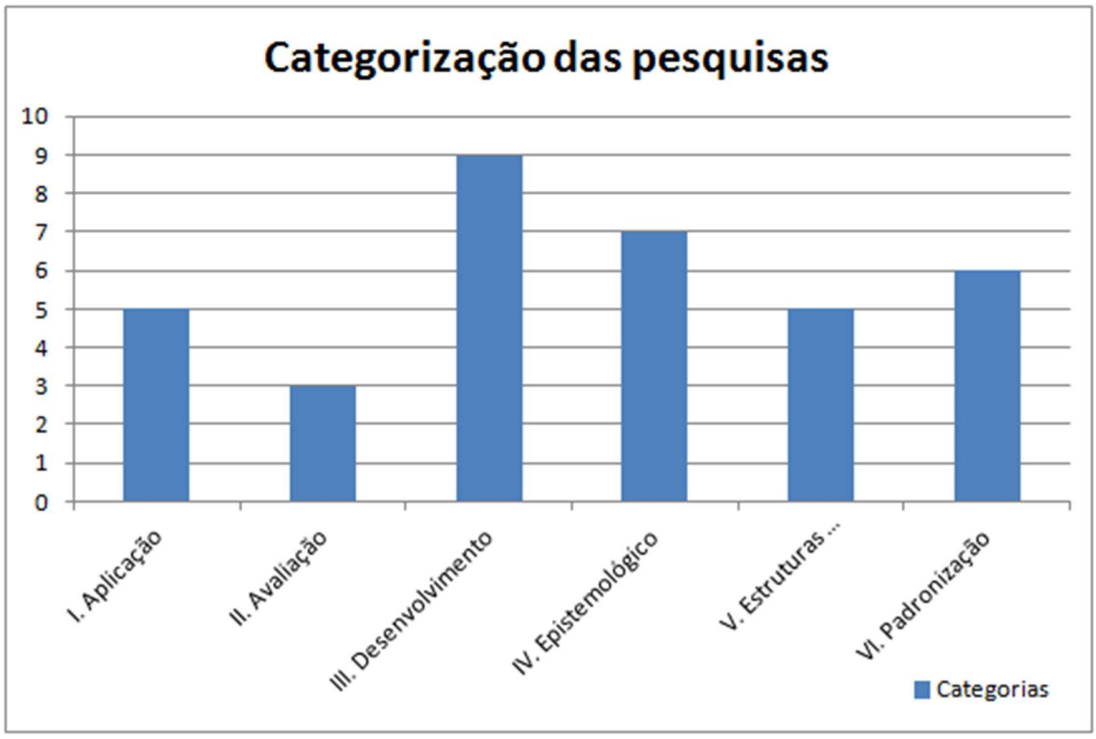
4 Análise das pesquisas por categorias

A partir da natureza das pesquisas, foram classificadas em cinco categorias:

I. **Aplicação** - Propostas ou aplicações de vocabulários controlados, tesauros, ontologias, etc. em ambientes digitais e instituições.
II. **Avaliação** - Estudos que buscam avaliar a eficácia e eficiência de vocabulários controlados em sistemas.
III. **Desenvolvimento** - Propostas de desenvolvimento de vocabulários controlados, ontologias, tesauros, etc. em ambientes digitais.
IV. **Epistemológico** - Pesquisas que abordaram e refletiram questões gerando conceitos, fundamentações, técnicas, estudos, explanações e críticas a cerca dos Sistemas de Organização do Conhecimento.
V. **Estruturas computacionais** - Trabalhos que apresentaram implementação de linguagens computacionais como RDF, OWL, etc. em vocabulários controlados, tesauros, etc.
VI. **Padronização** - Iniciativas de padronização de vocabulários controlados, terminologias, ontologias, tesauros, etc.

A figura 1 representa o mapeamento das pesquisas classificadas nas categorias apresentadas anteriormente.

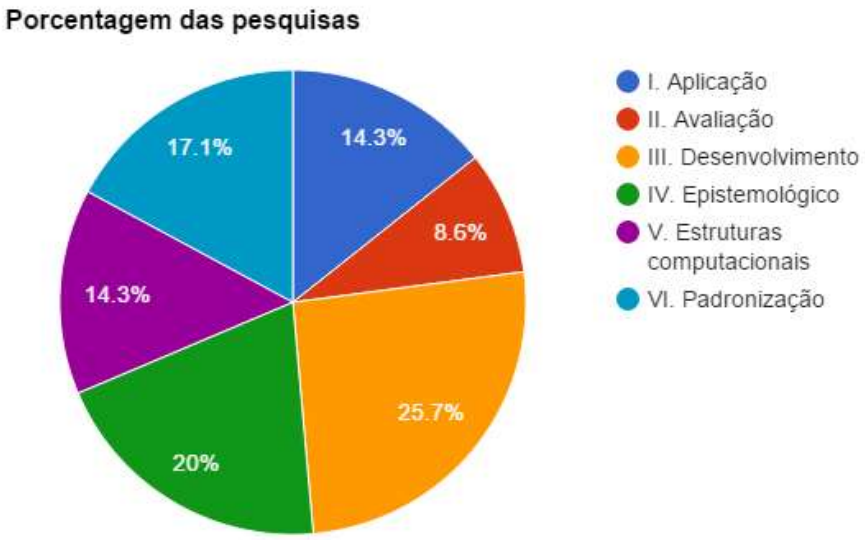
Figura 1. Frequência das pesquisas em relação as categorias



Fonte: Elaboração dos autores.

Conforme a figura 1, na categoria I. Aplicação foram classificados cinco (5) trabalhos, representando 14,3% das pesquisas; a categoria II. Avaliação, apresentou três (3) pesquisas, ou seja, 8,6%; na categoria III. Desenvolvimento destaca-se com nove (9) pesquisas, totalizando 25,7%; na categoria IV. Epistemológico apresenta-se 20% das pesquisas, ou seja, sete (7) trabalhos, já a categoria V. Estruturas computacionais apresentaram cinco (5) trabalhos, ou 14,3%, e por fim, a categoria VI. Padronização, com seis (6) trabalhos apresentados, representando 17,1% das pesquisas. A figura 2, apresenta a distribuição das porcentagens das pesquisas.

Figura 2. Distribuição por porcentagem das pesquisas



Fonte: Elaboração dos autores

Conforme pode ser observado, as pesquisas apresentadas nos eventos do DCMI que abordaram Sistemas de Organização do Conhecimento têm a natureza maior em propostas de seu desenvolvimento (categoria III.), em segundo lugar os estudos epistemológicos (categoria IV.), em terceiro lugar as questões envolvendo a padronização (categoria VI.), seguido das aplicações empatada com estrutura (categorias I. e V. respectivamente) e por último referente a avaliação (categoria II.).

5 Considerações

A partir da análise das pesquisas apresentadas no evento do DCMI, observa-se que o panorama dos estudos em Sistemas de Organização do Conhecimento pode ser dividido em 6 categorias: aplicação, avaliação, desenvolvimento, estudos epistemológicos, estruturas computacionais e padronização. O estado da arte revelou que a maior parte dos estudos estão voltados para construção e desenvolvimento de tesauros, ontologias e vocabulários controlados, categoria 3. Seguido dos estudos epistemológicos, que somados, contabilizaram mais de 45% das pesquisas apresentadas. Poucas pesquisas estavam relacionadas a avaliação de Sistemas de Organização do Conhecimento com aproximadamente 8,6 % das pesquisas.

Sugere-se como pesquisas futuras, a ampliação do recorte da pesquisa nos resumos apresentados nas conferências DCMI, assim como, a aplicação e comparação do estudo em outros eventos da área e periódicos. Essa análise poderá ampliar o panorama e revelar quais as tendências dos estudos em Sistemas de Organização do Conhecimento.

Referências

AMBIAH, N.; LUKOSE, D. Enriching webpages with semantic information. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 12., 2012, Sarawak. **Anais**… Malásia: DCMI, 2012. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3681>. Acesso em: 26 abr. 2015.

BARTOLO, L. et al. Development of a wiki-based, expert community-driven nanosystem vocabulary. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. **Anais**… México: DCMI, 2006. Disponível em: < http://dcpapers.dublincore.org/pubs/article/view/852>. Acesso em: 16 abr. 2015.

BOSCH, T. et al. Leveraging the DDI model for linked statistical data in the social, behavioral, and economic sciences. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 12., 2012, Sarawak. **Anais**… Malásia: DCMI, 2012. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3654>. Acesso em: 23 abr. 2015.

CAMPBELL, D. Identifying the identifiers. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 7., 2007, Cingapura. **Anais**… Cingapura: DCMI, 2007. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/868>. Acesso em: 20 abr. 2015.

CARLAN, E.; MEDEIROS, M. B. B. Sistemas de Organização do Conhecimento na visão da Ciência da Informação.
RICI: R.Ibero-amer. Ci. Inf., Brasília, v. 4, n. 2, p. 53-73, ago./dez.2011.Disponível em: <http://periodicos.unb.br/index.php/RICI/article/view/6209>. Acesso em 06 maio 2015.

CATARINO, M.; BAPTISTA, A. Relating folksonomies with Dublin Core. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. **Anais**… Alemanha: DCMI, 2008. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/915>. Acesso em: 21 abr. 2015.

CHEN, M.; LIU, X.; QIN, J. Semantic relation extraction from socially-generated tags: a methodology for metadata generation. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. **Anais**… Alemanha: DCMI, 2008. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/924>. Acesso em: 21 abr. 2015.

CURRIER, S. et al. Vocabularies for describing pedagogical approach in e-Learning: a scoping study. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA

APPLICATIONS, 6., 2006, Colima. **Anais**… México: DCMI, 2006. Disponível em: < http://dcpapers.dublincore.org/pubs/article/view/847>. Acesso em: 20 abr. 2015.

FRANCESCONI, E.; PERUGINELL, G. Retrieval of Italian legal literature: a case of semantic search using legal vocabulary. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. **Anais**… Espanha: DCMI, 2005. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/807>. Acesso em: 15 abr. 2015.

FREIRE, N.; BORBINHA, J.; CALADO, P. A language independent approach for aligning subject heading systems with geographic ontologies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Hague. **Anais**… Holanda: DCMI, 2011. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3624>. Acesso em: 26 abr. 2015.

GREENBERG, J. et al. Growing vocabularies for plant identification and scientific learning. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. **Anais**… Espanha: DCMI, 2005. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/808>. Acesso em: 20 abr. 2015.

HARPER, C. Encoding Library of Congress Subject Headings in SKOS: Authority control for the Semantic Web. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. **Anais**… México: DCMI, 2006. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/842>. Acesso em: 21 abr. 2015.

HILLMANN, D.; DUNSIRE, G.; PHIPPS, J. Maps and gaps: strategies for vocabulary design and development. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 13., 2013, Lisboa. **Anais**… Portugual: DCMI, 2013. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3673>. Acesso em: 25 abr. 2015.

HONMA, T.; NAGAMORI, M.; SUGIMOTO, S. Find and combine vocabularies to design metadata application profiles using schema registries and LOD resources. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 13., 2013, Lisboa. **Anais**… Portugual: DCMI, 2013. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3675>. Acesso em: 25 abr. 2015.

JEONG, W. Is tagging effective?: overlapping ratios with other metadata fields. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS,9., 2009, Seul. **Anais**… Coréia do Sul: DCMI, 2009. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/975>. Acesso em: 19 abr. 2015.

KAKALI, C. et al. Integrating Dublin Core metadata for cultural heritage collections using ontologies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 7., 2007, Cingapura. **Anais**… Cingapura: DCMI, 2007. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/871>. Acesso em: 18 abr. 2015.

KIM, H. et al. The state of the art in tag ontologies: a semantic model for tagging and folksonomies. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. **Anais**… Alemanha: DCMI, 2008. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/925>. Acesso em: 18 abr. 2015.

KURTH, M.; NEHLER, G.; SILTERRA, R. Using controlled vocabularies to manage resource relationships: the KMODDL experience. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. **Anais**… Espanha: DCMI, 2005. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/811>. Acesso em: 15 abr. 2015.

LAUSER, B. et al. Comparing human and automatic thesaurus mapping approaches in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. **Anais**… Alemanha: DCMI, 2008. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/918>. Acesso em: 20 abr. 2015.

LEE, W.; SUGIMOTO, S. Toward core subject vocabularies for community-oriented subject gateways. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. **Anais**… Espanha: DCMI, 2005. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/769>. Acesso em: 21 abr. 2015.

LIANG, A. et al. From AGROVOC to the Agricultural Ontology Service / Concept Server: An OWL model for creating ontologies in the agricultural domain. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. **Anais**… México: DCMI, 2006. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/841>. Acesso em: 19 abr. 2015.

MARZAL, M.; RUIZ, M.; CERVERÓ, A. Development of a controlled vocabulary for learning objects' functional description in an educational repository. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. **Anais**… México: DCMI, 2006. Disponível em: < http://dcpapers.dublincore.org/pubs/article/view/857>. Acesso em: 19 abr. 2015.

MASON, J. Knowledge management and Dublin Core. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 9., 2009, Seul. **Anais**… Coréia do Sul: DCMI, 2009. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/953>. Acesso em: 21 abr. 2015.

MENGER, M.; RüTHER, M. The Semantic Network Service: supporting heterogeneous environmental information systems. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. **Anais**… México: DCMI, 2006. Disponível em: < http://dcpapers.dublincore.org/pubs/article/view/860>. Acesso em: 21 abr. 2015.

MILES, A. et al. SKOS Core: simple knowledge organisation for the web. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. **Anais**… Espanha: DCMI, 2005. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/768>. Acesso em: 21 abr. 2015.

MILES, A. SKOS: Requirements for standardization. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. **Anais**… México: DCMI, 2006. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/839>. Acesso em: 11 abr. 2015.

MIYAZAWA, A. Parallel writing in east asian languages and its representation in metadata in light of the DCMI Abstract Model. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE

AND METADATA APPLICATIONS, 7., 2007, Cingapura. **Anais**… Cingapura: DCMI, 2007. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/863>. Acesso em: 17 abr. 2015.

MONTIEL-PONSODA, E. et al. Style guidelines for naming and labeling ontologies in the multilingual Web. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Hague. **Anais**… Holanda: DCMI, 2011. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3626>. Acesso em: 20 abr. 2015.

MORSHED, A. et al. Thesaurus alignment for Linked Data publishing. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Hague. **Anais**… Holanda: DCMI, 2011. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3623>. Acesso em: 19 abr. 2015.

PANZER, M., ZENG, M. Modeling Classification Systems in SKOS: Some Challenges and Best-Practice Recommendations. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 9., 2009, Seul. **Anais**… Coréia do Sul: DCMI, 2009. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/974>. Acesso em: 13 abr. 2015.

PATTUELLI, M.; NORBERG, L. Identity matters: constructing social identities through ontology-based metadata. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 6., 2006, Colima. **Anais**… México: DCMI, 2006. Disponível em: < http://dcpapers.dublincore.org/pubs/article/view/853>. Acesso em: 13 abr. 2015.

PÉREZ AGÜERA, J.; JIMéNEZ, R. Thesauri managing and software agents: a proposed architecture. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 5., 2005, Madrid. **Anais**… Espanha: DCMI, 2005. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/810>. Acesso em: 16 mar. 2015.

SUMMERS, E. et al. LCSH, SKOS and Linked Data. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 8., 2008, Berlin. **Anais**… Alemanha: DCMI, 2008. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/916>. Acesso em: 11 abr. 2015.

TIAN, M. et al. Towards the representation of Chinese traditional music: a state of the art review of music metadata standards. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 13., 2013, Lisboa. **Anais**… Portugual: DCMI, 2013. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3672>. Acesso em: 13 abr. 2015.

ZAVALINA, O. Free-text collection-level subject metadata in large-scale digital libraries: a comparative content analysis. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 11., 2011, Hague. **Anais**… Holanda: DCMI, 2011. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3630>. Acesso em: 13 abr. 2015.

ŽUMER, M.; ZENG, M.; HLAVA, M. A domain model for describing and accessing KOS resources: report of processes in developing a KOS description metadata application profile. In: INTERNATIONAL CONFERENCE ON DUBLIN CORE AND METADATA APPLICATIONS, 12., 2012, Sarawak. **Anais**… Malásia: DCMI, 2012. Disponível em: <http://dcpapers.dublincore.org/pubs/article/view/3656>. Acesso em: 13 abr. 2015.

Modelos conceituais e representação descritiva da informação

Elisabete Gonçalves de Souza
Universidade Federal Fluminense
elisabetes.souza@gmail.com

Wellington Freire Cunha Costa
Universidade Federal Fluminense
freirew@yahoo.com.br

1 Introdução

Os metadados são essenciais para a descrição documental pelo fato de transformarem a informação em códigos auxiliando na sua representação, recuperação e acesso. Os repositórios digitais fazem uso dos metadados para descrever e tornar possível a busca e a recuperação da informação em linha, possibilitando a interoperabilidade das bases de dados.

De acordo com Marcondes (2005, p.96):

> Um dos maiores objetivos do uso de metadados no contexto da Web é permitir não só descrever documentos eletrônicos e informações em geral, possibilitando sua avaliação de relevância por usuários humanos, mas também permitir agenciar computadores e programas especiais, robôs e agentes de *software,* para que eles compreendam os metadados associados a documentos e possam então recuperá-los, avaliar sua relevância e manipulá-los com mais eficiência.

Para garantir a integridade das descrições e interoperabilidade dos sistemas é primordial o uso de normas e padrões. Para os usuários essa ação permite-lhes localizar e escolher dentre as várias expressões e manifestações de um recurso, aquela que melhor atenda às suas necessidades de informação. Para os gestores, significa compartilhar os registros, usar e reusar as informações já processadas contextualizando-as de acordo com a política de seu sistema de informação.

O Dublin Core (DC) é o padrão para a estruturação de metadados mais utilizado para a criação de repositórios e apresenta as seguintes características: a) flexibilidade –

os elementos que o compõem são opcionais, podendo ser exibidos aleatoriamente e se houver necessidade podem ser repetidos ou modificados por qualificadores; b) simplicidade – é de fácil manuseio, sendo autoexplicativo, permitindo que o próprio autor do documento faça a descrição sem a necessidade de intensos treinamentos; c) extensibilidade – por ser um modelo simples e flexível, permite que diversas comunidades em áreas distintas utilizem o padrão DC trocando informações e obtendo o acesso a elas; d) interoperabilidade – um modelo comum de descrição utilizado por diversas áreas proporciona aos usuários maior habilidade nas pesquisas; e) âmbito internacional – embora tenha sido desenvolvido em inglês, já existem versões em outras línguas; infraestrutura adequada e escopo internacional (DCMI, 2015).

Conforme Marcondes (2005) o padrão DC foi pensado para ser simples o suficiente e autoexplicativo de modo que o próprio autor do documento possa descrevê-lo ao publicá-lo eletronicamente, o que seria impossível em outro padrão como o MARC, por exemplo, pois exige o aporte de profissionais de informação treinados dada a sua complexidade. Indo em direção oposta,

> o Dublin Core, pode ser definido como sendo o conjunto de elementos de metadados planejado para facilitar a descrição de recursos eletrônicos. A expectativa é de que os autores e Websites, que não possuam conhecimentos em catalogação, possuam capacidade de usar o Dublin Core para descrição de recursos eletrônicos, tornando suas produções mais visíveis aos mecanismos de busca e sistemas de recuperação (SOUZA; VENDRUSCULO; MELO, 2000, p. 93).

O DC é um padrão que complementa os demais, pela sua simplicidade na descrição dos registros, funcionando como um “meio-termo” entre dois tipos de descrição: as automáticas realizadas pelos Serviços de Indexação e Resumo e as catalográficas produzidas pela biblioteca. Por ser baseado no protocolo Z39.50 (hoje reestruturado como SRU - *Search and Retrieve URL,* e SRW - *Search/Retrieve Web Service*), permite a troca de informações entre diferentes sistemas de bases de dados, busca simultânea e importação de registros, podendo ser implantado em qualquer plataforma.

A interoperabilidade das descrições feitas no DC está garantida pelo uso da linguagem XML (*Extensible Markup Language*) e da semântica da RDF (*Resource*

Description Framework). De acordo com Marcondes (2005, p. 108) “enquanto a XML é uma linguagem genérica para a descrição da estruturação de documentos eletrônicos, RDF é específica para criar metadados com a finalidade de localizar e identificar recursos. A codificação em RDF permite descrever e pesquisar recursos de forma mais flexível, mediante representação de relacionamentos entre estes, na forma de triplas compostas por objeto-atributo-valor, sendo que um valor pode ser um literal ou outro recurso” (CAMPOS; CAMPOS; CAMPOS, 2005, p. 66). Ou seja: por meio da interoperabilidade semântica, o formato DC pode ser utilizado por diversas áreas e por ser flexível permite a extensibilidade e modularidade dos metadados, tornando possível a troca de informações entre as comunidades científicas de forma fácil e rápida.

2 Modelo entidade-relacionamento

O novo modelo conceitual FRBR (*Requisitos Funcionais para Registros Bibliográficos*) foi inspirado no metamodelo Entidade-Relacionamento (E-R) criado na década de 1970 pelo cientista da computação Peter Chen para otimizar as rotinas dos bancos de dados de modo que os registros gerados nos bancos fossem interligados, apurando a recuperação de itens importantes de forma fácil e rápida. Segundo Moreno (2006, p.34):

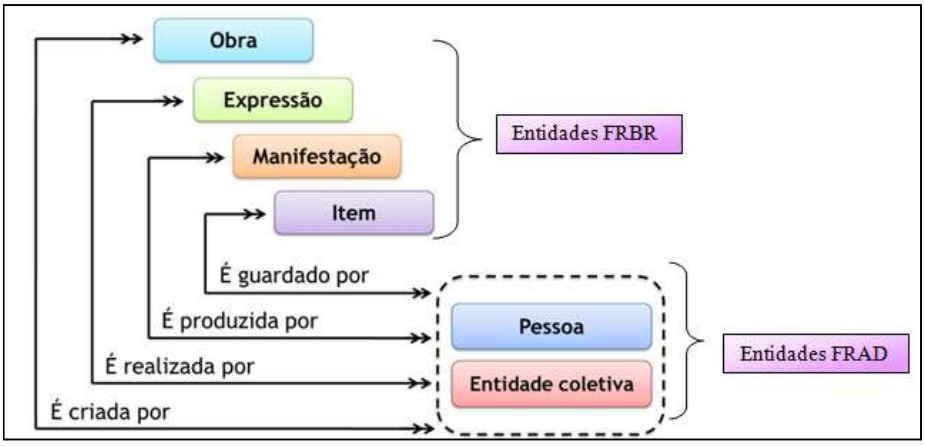
> Modelos de dados são um conjunto de conceitos utilizados para descrever um banco de dados. O modelo E-R é um modelo lógico com base em objetos, e a identificação de entidades e relacionamentos é entendida como a captura da semântica dos dados, para projetar um banco de dados. Este modelo não visa à implementação e sim à modelagem/representação dos dados. A partir desta modelagem, é possível implementar um banco de dados em outros modelos de dados: orientado a objeto, relacional, etc.

Segundo Silva e Santos (2012) a modelagem conceitual descreve que um domínio é composto por coisas representadas pelas entidades, que por sua vez possuem características a serem descritas, denominadas atributos; e pelos relacionamentos existentes entre elas. Por isso as bases de dados relacionais possuem maior capacidade semântica.

A figura abaixo ilustra o modelo E-R em um banco de dados com informações sobre um conjunto de objetos classificados como pessoa (entidade FRAD), um conjunto

de objetos classificados como manifestação (entidade FRBR), um conjunto de associações, cada uma ligando uma manifestação a uma pessoa – criada por (relacionamento) e um conjunto de informações que caracterizam as entidades (atributos).

Figura 1 – Modelo E-R (domínio bibliográfico)



Fonte: Assumpção (2012). Adaptação dos autores

O modelo (E-R) foi criado para melhorar o desenvolvimento dos bancos de dados fornecendo buscas mais precisas para os usuários. O modelo conceitual FRBR foi criado pela IFLA (*International Federation of Library Associations and Institutions*) também com esse propósito: representar as necessidades do usuário de catálogos e bibliografias levando em consideração suas tarefas ou funções - encontrar, identificar, selecionar e obter - com o objetivo de aproximá-lo da obra.

O modelo FRAD (*Functional Requirements for Authority Data*) é uma extensão dos FRBR para registro de autoridade, corresponde ao cabeçalho autorizado para uma entidade estabelecido de acordo com as regras de uma agência catalogadora. No universo bibliográfico uma entidade FRAD arrola informações sobre uma pessoa ou entidade, a qual tem seu nome usado como ponto de acesso controlado nos registros bibliográficos, seja em um catálogo ou em um repositório.

As entidades FRAD são responsáveis pela criação, produção, edição e outras responsabilidades relacionadas a uma obra. Seu uso amplia não só a capacidade de relacionamentos como a de associações de dados, já que ao individualizarmos uma entidade pela aplicação de atributos, podemos associá-la a dados abertos, como

currículo lattes, grupos de pesquisa, etc. principalmente para autores acadêmicos com produção em repositórios digitais.

2.1 Aplicação do modelo FRAD em repositório

O uso do padrão DC associado ao modelo E-R possibilitará que o usuário/autor ao depositar um documento possa descrevê-lo de forma mais precisa, já que o novo padrão RDA (*Resource Description and Access*) permite a descrição dos metadados bibliográficos e de autoridade de modo extensivo, em linguagem próxima daquela usada pelos usuários ao definir as entidades e seus atributos.

Para exemplificar usamos como campo empírico o repositório institucional LUME da Universidade Federal do Rio Grande do Sul. Para iniciarmos o processo de modelagem desse repositório contextualizamos sua estrutura aproximando-a das entidades da família FRBR. Para efeito de estudo, faremos a modelagem das entidades FRAD, acrescentando atributos que identificam e qualificam as pessoas responsáveis pela produção da obra (uma dissertação de mestrado) e a entidade/instituição de pesquisa com a qual se relaciona.

Figura 2 - Identificação das entidades

Fonte: elaboração dos autores

A simulação abaixo demonstra a modelagem da estrutura descritiva do Repositório Institucional (RI) LUME, campos referentes aos dados de autoridade. A adição de atributos FRAD aos elementos DC contributor.advisor e contributor.author que

representam pontos de acesso/pessoa da obra "*Gestor de quem?" A contribuição do bibliotecário para a gestão de pessoas da organização : um estudo de caso"* visa aprimorar processo de busca e recuperação da informação no sentido que permite aos usuários do repositório novas possibilidades de relacionamentos durante a pesquisa. Com isso, o usuário terá mais informações sobre o criador do documento/recurso e os diferentes relacionamentos que este mantém com outras entidades, no caso de nosso exemplo, seu orientador, o programa no qual está filiado, grupos de pesquisa, etc.

Nessa direção, podemos dizer que o acréscimo de novos atributos (currículo lattes, filiações institucionais, etc.) aos metadados de autoridade identificados no registro DC (autor, orientador) amplia os processos de busca e recuperação de informação; levam o pesquisador a navegar para além do conteúdo da obra à medida que a contextualizava em relação aos seus pares. Para recursos como as dissertações e teses, tais relacionamentos são fundamentais para a interlocução entre pesquisadores o que faz dos repositórios uma poderosa ferramenta de comunicação científica.

No caso do repositório LUME, para que as relações fossem mais completas seria interessante a descrição dos metadados dos membros da banca. No entanto, essa informação foi suprimida, porque o gestor, a Universidade Federal do Rio Grande do Sul, segue para esse tipo de obra (teses e dissertações) os metadados estruturais (MTD-BR) definidos pela BDTD (Biblioteca Digital Brasileira de Teses e Dissertações), pois integram esse consórcio coordenado pelo IBICT.

A figura abaixo mostra a aplicação do modelo conceitual FRAD na estrutura descritiva do RI-LUME referente aos dados de autoridade (entidade pessoa) com o acréscimo de atributos que identificam as autoridades autor e orientador de forma contextualizada, no que diz respeito à sua inserção acadêmica. Além da forma extensível**,** adotamos também cabeçalhos para as entidades relacionadas aos nomes**,** por ser esta a forma normalizada predominante na maioria dos catálogos, o que nos permite interligar os dados sobre as diferentes comunidades científicas.

Figura 3 – Aplicação do FRAD no RI-LUME

Entidade FRAD	dc.contributor.advisor	Nakayama, Marina Keiko
Atributo FRAD	Link Curiculo Lattes	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4798728D2
Atributo FRAD	Linha de pesquisa	Gestão do conhecimento e compartilhamento
Atributo FRAD	Grupo de pesquisa	Núcleo Estudos e Observação de Gestão, Aprendizagem e Pessoas
Atributo FRAD	Link grupo/pesquisa	http://www.neogap.ufsc.br/wp/?page_id=58
Entidade FRAD	dc.contributor.author	Soares, Monica Fonseca
Atributo FRAD	Link Currículo Lattes	http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=K4782635D0
Atributo FRAD	Programa/Pós-graduação	Programa de Pós-Graduação em Administração
Atributo FRAD	Forma normalizada do nome do Programa	Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Administração
Atributo FRAD	Link para o programa	http://www.ufrgs.br/escoladeadministracao/pos-graduacao/

Fonte: elaboração dos autores

Na institucionalização de um campo científico é necessário que se tenha confiabilidade no conhecimento produzido. Um repositório institucional é um meio de divulgação do conhecimento e resultados advindos das pesquisas empreendidas pelos cientistas; um canal de comunicação que arrola diferentes produções formais e informais cujos resultados obtidos podem ser acessados e compartilhados por outros pesquisadores. Por isso**,** para além da flexibilidade proporcionada pela aplicação do modelo FRBR, cabe ao gestor acuidade quanto à descrição e codificação dos metadados, levando em consideração o uso da RDA e da linguagem RDF, aferindo assim, as vantagens advindas do uso desses padrões na recuperação e acesso a informações em ambiente digitais.

3 Considerações parciais

Como vimos, um repositório institucional tem o objetivo de preservar o conhecimento acumulado nas instituições apresentando como principais vantagens o armazenamento de grande quantidade de informações, facilidade de busca e recuperação e autonomia para o produtor fazer a descrição e o autoarquivamento de seus documentos.

O processo de simulação da aplicação dos atributos do FRAD na estruturação descritiva, demonstrou que tal implementação melhorará consideravelmente o processo de busca e recuperação da informação, qualificando os registros com o acréscimo de atributos individuais como o currículo lattes e de contextualização acadêmica, como a vinculação aos grupos e/ou núcleos de pesquisa, o que permitirá ao usuário uma visão

mais abrangente sobre as comunidades científicas nas quais se inserem os autores e seus pares.

No que diz respeito à recuperação da informação dois aspectos merecem destaques: a) o detalhamento de um registro com acréscimos de atributos amplia as possibilidades de relacionamentos entre as entidades pessoas e as demais entidades bibliográficas que compõem o registro DC; b) a incorporação de links (URL) como o currículo lattes inaugura a possibilidade de associar as ferramentas bibliográficas aos dados abertos disponíveis na web.

Referências

ASSUMPÇÃO, Fabrício. **O que é FRBR?** Disponível em:<http://fabricioassumpcao.com/2012/07/o-que-e-frbr.html>. Acesso em: 4 maio 2015.
.

CAMPOS, M. Luiza Machado; CAMPOS, M. Luiza de Almeida; CAMPOS, Linair Maria. Web semântica e a gestão de conteúdos informacionais. In: MARCONDES, Carlos Henrique et al. (Org.). **Bibliotecas Digitais:** saberes e práticas. Salvador, EDUFBA; Brasília; IBICT, 2005. p. 55-74.

DCMI. **Using Dublin Core.** Disponível em: http://dublincore.org/documents/usageguide/ Acesso 30 abr. 2015.

MARCONDES, Carlos Henrique. Metadados: descrição e recuperação na Web In: ____ et al. (Org.). **Bibliotecas digitais:** saberes e práticas. Salvador, BA: EDUFBA; Brasília; IBICT, 2005. p. 77-143

MORENO, Fernanda Passini. **Requisitos funcionais para registros bibliográficos – FRBR**: um estudo no catálogo da Rede Bibliodata. 2006. 199 f. Dissertação (Mestrado em Ciência da Informação) – Universidade de Brasília, 2006.

SILVA, Renata Eleutério da; SANTOS, Plácida Leopoldina Ventura Amorim da Costa. Requisitos Funcionais para Registros Bibliográficos (FRBR): considerações sobre o modelo e sua implementabilidade. **RBBD**. São Paulo, v. 8, n. 2, p.116-129, Jul./Dez. 2012.

SOUZA, M. I. F.; VENDRUSCULO, L. G.; MELO, G. C. Metadados para a descrição de recursos de informação eletrônica: utilização do padrão Dublin Core. **Ciência da Informação**, v. 29, n. 1, p. 93-102, Jan./Abr. 2000.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. **Repositório Lume**. 2015. Disponível em: http://www.lume.ufrgs.br Acesso 30 abr. 2015.

Modelagem de nanopublicações: abordagem experimental

Lorena Tavares de Paula
Universidade Federal de Minas Gerais
lorena.ltp@gmail.com

Maria Aparecida Moura
Universidade Federal de Minas Gerais
cidamoura@gmail.com

1 Introdução

A organização da informação em meio digital requer a ampliação dos fundamentos teóricos e das metodologias para o tratamento de massas documentais. Para tanto, há a necessidade de desenvolver propostas de organização de informação que orientem a prática profissional considerando-se o novo cenário de produção documental.

Nesse sentido, o presente artigo destaca um experimento metodológico para a modelagem de nanopublicações em interface com a indexação em ambientes digitais.

2 Nanopublicação indexação: aspectos conceituais

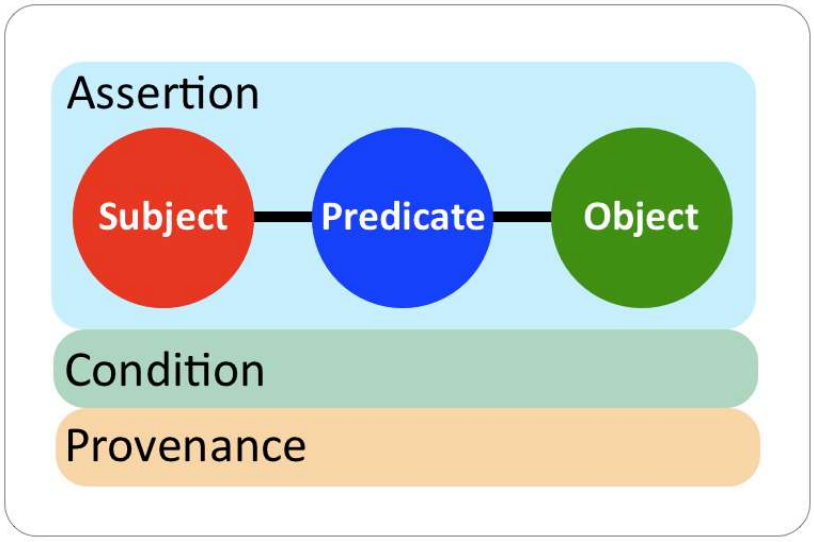
A proposta de elaboração de nanopublicações, estabelecida pela *Concept Web Alliance* (CWA), objetiva o desenvolvimento de sistemas que delimitem padrões para a representação de informações científicas. Em uma descrição simplificada, é possível afirmar que as nanopublicações efetivam-se através do mapeamento de conceitos dispostos em publicações digitais em um dado domínio de conhecimento.

A nanopublicação possibilita uma comunicação que se baseia em informações contextuais de alta qualidade e apoia a divulgação, a apropriação e a organização de informações contextualizadas de acordo com um autor e uma unidade conceitual. Groth, Gibson, Velterop (2011) explicam que a elaboração de uma nanopublicação requer conteúdos estruturais

que reforcem o valor de uma declaração ao ponto em que ela possa ser considerada uma publicação.

A nanopublicação é, pois, uma aplicação computacional para a representação de afirmações conceituais em contexto.

Figura 1: Anatomia da nanopublicação



Fonte: nanopub.org

De acordo com a FIGURA 1, a "anatomia da nanopublicação" é caracterizada por uma afirmação (*Assertion*) assinalada pela relação semântica estabelecida pelo "sujeito, o predicado e o objeto". Também possui declarações (*Conditions*) condicionadas à identificação da afirmação. Além disso, é caracterizada por proveniências (*Provenance*) que estabelecem a identificação dos elementos, os quais legitimam a nanopublicação, como: autor, localização (URL), vínculos institucionais, dentre outros. Os elementos "*assertion"; "conditions" e "provenance"* são materializadas em aplicações computacionais que utilizam padrões abertos, como o XML (*Extensible Markup Language*), e RDF (*Resource Description Framework*), para fomentar a ação colaborativa de integração de dados.

Em seu aspecto conceitual, as nanopublicações devem ser citáveis, atribuíveis e revisáveis. Além disso, elas precisam ser reorganizáveis, pois devem ser facilmente agregadas e identificadas em toda a web. Segundo Groth, Gibson e Velterop (2011), elas precisam ser extensíveis para atender as novas formas de descrição orientadas por conjuntos de definições.

Dessa forma, as nanopublicações estabelecem uma representação temática do documento, por meio da evidência conceitual e da afirmação de autorias e comunidades de conhecimento.

No campo da organização da informação, o processo de indexação objetiva estabelecer o assunto (aboutness) de um documento obedecendo a dois princípios básicos:

1 - Análise Conceitual;

2 - Tradução para uma linguagem de indexação.

A análise conceitual implica decidir do que trata o documento. Uma indexação de assunto eficiente sugere que se tome uma decisão não somente quanto ao que é tratado no documento, mas também quanto ao porquê ele se reveste em provável interesse temático para determinado grupo de usuários. Já a fase de tradução para uma linguagem de indexação relaciona-se à necessidade de controle de vocabulário, tendo em vista melhorar o processo de recuperação de informação.

Vale dizer que todo o processo de indexação objetiva melhorar a representação específica de um documento e a sua consequente recuperação, a qual Cesarino (1985, p.85) descreve como um **"**conjunto de cooperações consecutivas executadas para localizar, dentro da totalidade de informações disponíveis, aquelas realmente relevantes". A partir da descrição de assunto em forma de termos descritores, pode-se promover uma integração entre documento, acervo e usuário.

A nanopublicação visa, então, proporcionar uma representação de conteúdos, por meio da estruturação de afirmações conceituais contidas em uma publicação formal. Diante disso, apresenta-se uma metodologia experimental que objetiva integrar o processo de indexação à modelagem de nanopublicações derivadas de documentos eletrônicos.

3 Metodologia para a modelagem de nanopublicações

A metodologia para modelagem de nanopublicação incorpora os princípios de indexação, bem como três dimensões essenciais à representação da informação e do

conhecimento em contextos digitais: a dimensão informacional, a dimensão computacional e a interface.

No método científico, um experimento é um conjunto de ações a respeito de fenômenos físicos. O experimento se fundamenta numa abordagem empírica. Dependendo da perspectiva, ele poderá conduzir a uma compreensão objetiva e completa do mundo físico ou apenas a uma ajuda na melhoria do conhecimento subjetivo, através das apreensões teórico/metodológica.

Segundo Gil (2002, p. 69), o experimento pode ser considerado um método hipotético-dedutivo, de modo que, para o autor, “o cientista, através de uma combinação de observação cuidadosa e de hábeis antecipações, alcança um conjunto de postulados que governam os fenômenos pelos quais está interessado”. Os experimentos possuem, portanto, grande aceitação, sobretudo nas ciências naturais. Gil (2002, p. 73), ainda, afirma que, “nos círculos neopositivistas chega mesmo a ser considerado como o único método rigorosamente lógico”. Entretanto o autor também salienta que “sua aplicação às ciências sociais apresenta sérias limitações”. “Ainda assim, muitos autores o consideram suficiente para a construção de modelo lógicos de investigação em ciências sociais”.

No que diz respeito à Ciência da Informação, acredita-se que, como uma ciência social aplicada, ela pode oferecer ao campo métodos que conciliem a integração de processos para efetivação da experimentação como método de pesquisa. Por mais que o contexto social possa sofrer adaptações e alterações, um corpus documental selecionado para uma experimentação de modelagem pouco sofrerá com as mudanças do contexto sociocultural. Nesse sentido, os experimentos propostos a partir desse experimento metodológico podem ser compreendidos como uma simulação.

> A simulação é um "experimento virtual" que requer um modelo operatório representando no todo (ou em parte) um sistema ou processos que o caracterizem. O uso dessa metodologia é analisado como forma de buscar um modelo, confirmá-lo ou fazer uma projeção de eventos futuros. Dessa forma, a simulação se presta tanto no "contexto de descoberta" quanto no "contexto da prova". (VICENTE, 1995, on-line)

A simulação é, dessa maneira, "a construção e a manipulação de um modelo operatório representando todo, ou parte de um sistema ou processos que o caracterizam".

A simulação funciona como um "modelo refletindo as características centrais de um sistema, processo ou ambiente, real ou proposto" (KLEIBOER, 1997, p. 18).

A metodologia para desenvolvimento dos experimentos de modelagem de nanopublicações considera a hipótese de que a nanopublicação pode ser um modelo de representação de informação em ambiente digital, quando colocada em interface com os princípios de indexação.

No que diz respeito a essas considerações, são descritas, nas seções a seguir, as fases da metodologia, o experimento de modelagem.

3.1 Fase 1 – Geração de mapeamento do corpus a ser pesquisado: dimensão informacional

Em uma primeira etapa, é necessário selecionar periódicos oriundos de campos de conhecimento correlativos e que utilizam e disseminam o conceito a ser modelado em formato de nanopublicação. Para todas as entradas de um dado conceito, é possível extrair os seguintes metadados: título do artigo, título da revista, volume, número, página inicial do artigo, data de publicação, nomes e sobrenomes de autores, URL.

Quadro 1 – Elementos de identificação do corpus “fase 1” (proveniência da nanopublicação)

Conceito	
Título do artigo **Título da revista**	
Número, volume	
Página inicial do artigo que consta o conceito	
Data da publicação	
Nome do autor **URI - (*Uniform Resource Identifier*).**	

Fonte: Elaborado pelas autoras, 2015.

O QUADRO 1 apresenta os metadados para estabelecer uma estrutura que permite consultar, identificar e localizar as informações conceituais no corpo dos artigos. O plano de extração e identificação dos conceitos que poderão compor a nanopublicação dependerá diretamente das informações consolidadas a partir do quadro descrito na fase 1. Essas informações materializarão as e-evidências das nanopublicações.

Como exemplo pode-se considerar a intenção de modelagem em nanopublicação do conceito “inovação”. O primeiro passo seria pesquisar o termo “inovação” em uma base de dados, como a Scielo, na qual se obteria resultados de ocorrência desse conceito em artigos de periódicos de campos de conhecimento distintos. O corpus a ser formado para representar a modelagem do conceito “inovação” deveria ser descrito seguindo as orientações do quadro descrito na fase 1.

3.2 Fase 2 – Extração conceitual a partir dos princípios de indexação: dimensão informacional

Os conceitos a serem modelados em nanopublicações podem ser gerais ou individuais e revelarem a complexidade própria ao domínio ao qual estão associados.

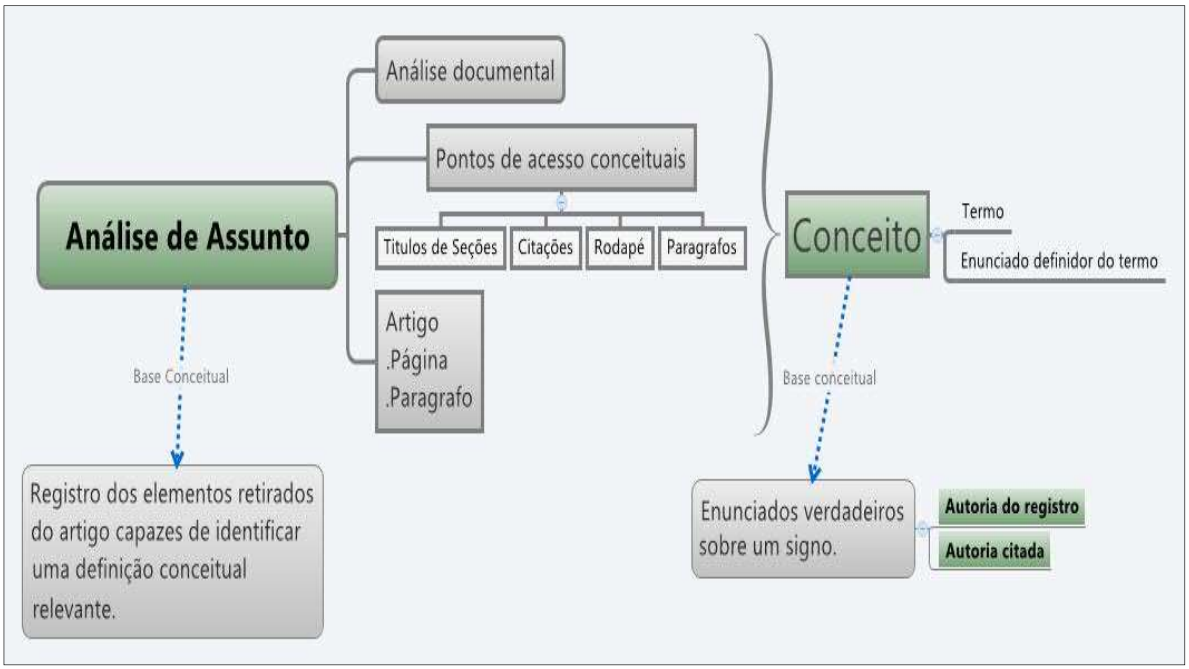
HJORLAND (2001, p. 774), citando Wilson (1968), destaca os elementos mais importantes na determinação do assunto de um documento para a extração de conceitos relevantes:

a) identificar o propósito do autor ao escrever o documento;
b) Identificar o domínio e subordinação dos diferentes elementos conceituais;
c) observar grupo de conceitos e referências estabelecidos;
d) estabelecer um conjunto de regras de seleção para os elementos "essenciais" (em contraste com o não essencial) do documento em sua totalidade.

Para Hjorland (2001, p. 774), o método para determinar o assunto de um documento está intimamente relacionado com as teorias de significado, interpretação e

epistemologia. Nesse sentido, o processo de extração conceitual proposto nesta metodologia segue as perspectivas da identificação de atinência (assunto) proposta por Hjorland.

Figura 2 – Extração conceitual



Fonte: Elaborado pelas autoras, 2015.

De acordo com a FIGURA 2, faz-se a análise de assunto de cada artigo que compõe o corpus. Nesse processo selecionam-se os conceitos que possuem definições explícitas e auxiliam na identificação do domínio de conhecimento evidenciado. Esses conceitos podem ser extraídos de: títulos de seções, citações, rodapés e parágrafos. Posteriormente, identifica-se o conceito através de um “termo” e um “enunciado” que estabelece a definição nominal.

Seguindo o exemplo com a palavra “inovação”, serão identificados, por meio da análise de assunto, os conceitos fundamentais que integram a formulação da definição de “inovação”, sempre observando as diferentes áreas de conhecimento em que o conceito é formulado.

Nessa fase é essencial a marcação, no corpo do artigo, a página e o parágrafo que destacam o conceito. Em seguida fazem-se as devidas referências de autoria:

a) autores do registro;
b) autores citados para a formulação conceitual;
c) comunidade de conhecimento de origem (a qual o campo de conhecimento o autor do registro e o autor citado pertencem).
d) enunciados para a definição conceitual

A partir da identificação, caracterização e marcação dos conceitos, juntamente com suas definições e caracterização do domínio de conhecimento, pode-se iniciar a modelagem da nanopublicação.

3.3 Fase 3 – Modelagem da Nanopublicação: dimensão informacional

Seguindo as definições da fase 1, que estabelece os identificadores do corpus selecionado, e da fase 2 com o processo de identificação e seleção conceitual, espera-se estabelecer identificações, como:

a) identificações de definições conceituais (fase 2)
b) autorias representativas (fase2)
c) campos de conhecimentos nos quais atuam os autores e os conceitos (fase1)
d) localização das publicações (fase 1)

A partir da identificação desses elementos, pode-se iniciar a modelagem da nanopublicação seguindo as perspectivas da tripla sintaxe: sujeito – predicado - objeto.

Seguindo o exemplo com o conceito “inovação”, as relações de sintaxe serão estabelecidas na seguinte ordem: Sujeito = Inovação / predicado = conectores de ligação (verbo) / Objeto: formulação conceitual.

Figura 3 – Modelagem da nanopublicação

Fonte: Elaborado pelas autoras, 2015.

Na FIGURA 3, observamos que a primeira ação consiste na identificação do conceito. A adequada extração da definição conceitual é essencial para delimitar a compreensão e a reutilização das nanopublicações. Na definição, deve-se destacar a proveniência, de acordo com os processos delimitados na fase 1 (descrição do corpus).

Em sequência, demarcam-se os processos de tripla sintaxe: sujeito – predicado – objeto. Nesse momento, descreve-se a nanopublicação em suas relações conceituais para a construção de afirmações que oferecerão suporte para a declaração.

As declarações consistem em uma identificação textual relacionada com uma identificação autoral. Descreve-se a nanopublicação com uma abordagem orientada pela representação de dados estruturados, juntamente com a sua proveniência em uma entidade publicável e citável.

Diante disso, a nanopublicação consiste na afirmação, verificada na tripla sintaxe, e a

procedência da afirmação é materializada na declaração contextualizada com as autorias e comunidades de conhecimento que lidam com o conceito.

As anotações oferecem a completude dos assuntos mencionados pela nanopublicação. Esses assuntos têm a identificação do autor do registro e do autor citado nas referências do artigo, bem como das definições conceituais atribuídas.

Cabe ressaltar que esses elementos associados constituem a concretização da nanopublicação e apresentam mecanismos para garantir a integridade do conceito e a sua origem de divulgação juntamente com a responsabilidade autoral. Além disso, garantem o acesso à proveniência, permitindo aos usuários avaliarem a confiabilidade dos conceitos. Sobretudo, fornecem um mecanismo pelo qual os autores e as instituições podem ser reconhecidos por sua contribuição para formulação de um conceito em um dado campo científico.

Desse modo, as publicações listadas serão a URL de identificação das nanopublicações que oferecerão elementos para a aplicação computacional da nanopublicação. A combinação entre a afirmação e a procedência da afirmação, que materializa a origem da nanopublicação, representa a entidade publicável e citável.

3.4 Fase 4 – Aspectos tecnológicos - aplicação da Web Semântica na modelagem da nanopublicação: dimensão computacional

As nanopublicações podem expor dados quantitativos e qualitativos.

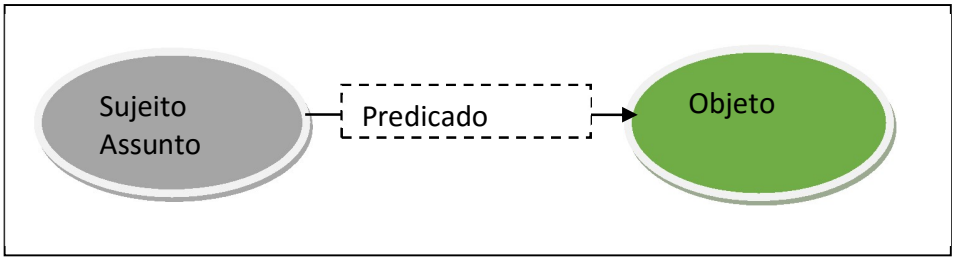
A representação computacional das nanopublicações pode ser consultada de maneira automatizada usando tecnologias de Web Semântica. A aplicação computacional orientada às nanopublicações baseia-se em RDF (*Resource Description Framework*)**.** A partir do processamento em RDF, podem-se estabelecer as representações da tripla sintaxe (Sujeito, predicado, objeto), característica da nanopublicação. Essa tecnologia sugere uma representação para o conteúdo web, que permite a integração das bases de dados e a análise de informações pelos aplicativos do software, tendo como base o *Uniform Resource Identifier* (URI).

A descrição das nanopublicações é feita essencialmente por declarações ou afirmações que, graficamente, denotam a ligação entre dois nós de um gráfico. Nessa

perspectiva, o primeiro nó é o Assunto (sujeito), enquanto o nó de destino é chamado Objeto. O segmento que conecta assunto (sujeito) e objeto é o predicado. Nessa modelagem todos os elementos são rotulados pelo respectivo URI.

A figura a seguir mostra um exemplo de um gráfico RDF com apenas uma afirmação.

Figura 12 – Modelo simplificado de RDF



Fonte: Elaborado pelas autoras, 2015.

A modelagem da nanopublicação em RDF pode seguir critérios sugeridos pelo nanopb.org (2015):

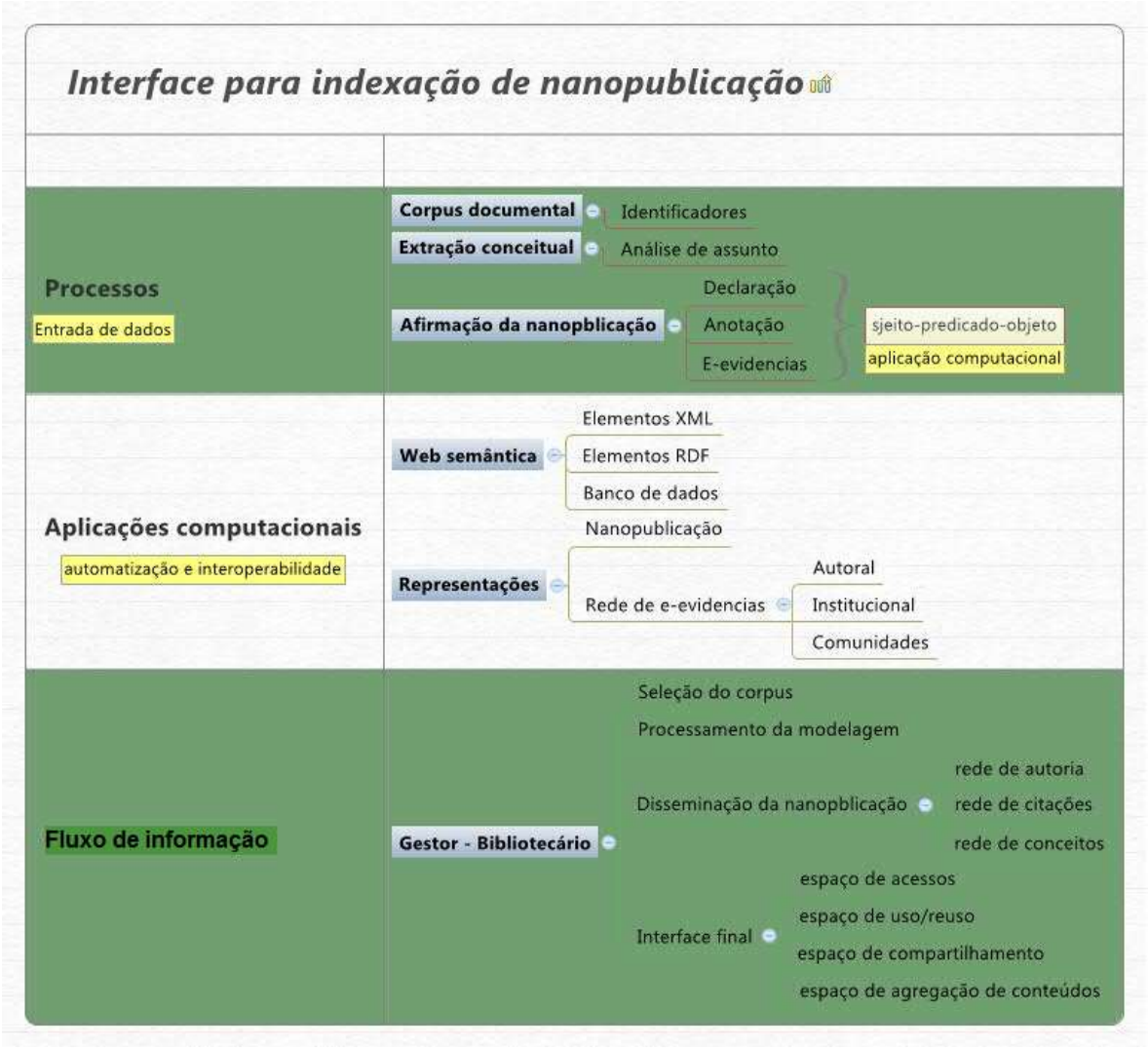
a) a nanopublicação será orientada por um conjunto de quatro RDF (sujeito-predicado-objeto + contexto);
b) o contexto (o gráfico) de cada triple tem de ser especificado como um URI válido;
c) explicitações de Declarações;
d) explicitações de Afirmações;
e) materialização de-E-evidências da nanopublicação.

O uso de RDF para representar a modelagem da nanopublicação justifica-se em seus processos descritivos simples. Além disso, o RDF possibilita apresentar e explorar o trabalho cooperativo (MANOLA; MILLER, 2002). Isso acontece por meio da autenticação de vários *namespaces* relacionando conceitos e autorias em comunidades de contextos distintos.

3.5. Fase 5 – Proposição de interface

Uma interface baseada nas tecnologias Web semântica deve seguir diretrizes que auxiliem o usuário a compreender as propriedades globais do sistema. Desse modo, é necessário propor elementos essenciais que caracterizarem as funcionalidades pretendidas no sistema para a organização da informação. Nessa concepção, podem-se sugerir a priori os elementos de organização de informação fundamentais para o sistema:

Figura 4 – Sugestão de interface



Fonte: Elaborado pelas autoras, 2015.

De acordo com a FIGURA 4, o “Sistema de Entrada” deve ser baseado nos processos de identificação do corpus e princípios de indexação que são apresentados

nas fases 1 (um) e 2 (dois) da metodologia. Todos os elementos de entrada, na qual o corpus selecionado é mapeado para posterior modelagem da nanopublicação, podem ser controlados pelo gestor.

O indexador de nanopublicações estabelece os processos de extração conceitual e as aplicações da tripla sintaxe na elaboração das declarações e afirmações da nanopublicação.

Salienta-se, contudo, que, nos processos orientados pela indexação, nem todos os aspectos da nanopublicação podem ser estabelecidos por processos automatizados das tecnologias da web semântica.

Em sequência, deve-se aplicar a camada tecnológica baseada na web semântica. Essa camada computacional irá proporcionar os ajustes técnicos necessários aos processos de automação, gerenciamento e disseminação das nanopublicações em consonância com a interoperabilidade. Por fim, no sistema de saída, revela-se uma rede de informações baseadas nas declarações, afirmações e e-evidencias das nanopublicações.

Na interface de aplicação computacional, por sua vez, verifica-se a necessidade de emprego do XML e RDF. Essas aplicações devem proporcionar uma infraestrutura computacional que contenha tanto um banco de dados relacional quanto a possibilidade de construção colaborativa de informações. Essas funções tecnológicas devem promover, na saída do sistema, a consolidação da declaração da nanopublicação e as orientações de identificação em forma de uma rede autoral e institucional tecida a partir dos elementos de e-evidência.

Ressalta-se, por fim, que o fluxo de informações observado no sistema de interface demonstra o papel do gestor em sua função de separar o corpus e aplicar as fases metodológicas essenciais aos experimentos. A partir dessa ação, tem-se a perspectiva de estabelecer nanopublicações e suas redes de autoria, conceitos e citações.

4 Considerações finais

Nos procedimentos clássicos de indexação, ocorre a representação da informação com termos que correspondem simultaneamente ao conteúdo intelectual do documento

e à provável questão de busca do usuário. Entretanto, diante do crescente volume de informações publicado em meio digital, pode se fazer necessária a seleção de “unidades conceituais” contextualizadas que auxiliem na construção de novos conhecimentos. Sugere-se que tais unidades possam ser consideradas “nanodocumentos”. Estes, por sua vez, são a menor unidade conceitual citável livre de ambiguidades e se fundamentam na afirmação de um autor realizada em publicações formais.

Tem-se, dessa maneira, a hipótese de que as nanopublicações podem ser consideradas um modelo emergente de representação da informação em ambientes digitais com orientações metodológicas estabelecidas em diálogo com a indexação. A confirmação disso poderá ser obtida através de experimentos e simulações que possuem, na metodologia sugerida neste artigo, os elementos norteadores fundamentais.

Nos ambientes digitais, portanto, percebe-se uma urgente demanda pela criação de mecanismos de representação de informação que consigam precisar um conteúdo informacional, sem distorções em relação ao contexto e ao entendimento conceitual. Nesse aspecto, a integração dos princípios de indexação com o processo de modelagem de nanopublicação pode ser considerada uma metodologia promissora.

Referências

CESARINO, M. A. N.; PINTO, M. C. M. F. Análise de assunto. **Revista de Biblioteconomia de Brasília,** Brasília, v. 8, n. 1, p. 32-43, jan./jun. 1985.

DAHLBERG, Ingetraut. Teoria do conceito. **Ciência da Informação**, Brasília, v. 7, n.2, p. 101-107, 1978.

GIL, Antônio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002. 175 p.

GOMES, Hagar Espanha; GUSMÃO, Heloísa Rios. **Guia prático para elaboração de índice**. Niterói: APB-RJ, 1983.

GROTH, P; GIBSON.A; VELTEROP, J. **The Anatomy of a Nano-publication**. Disponível em <http://www.w3.org/wiki/images/c/c0/hclsig$$swansioc$$actions$ $rhetoricalstructure$$meetings$$20100215$cwa-anatomy-nanopub-v3.pdf> Acesso em: 14 mar. 2014.

HJORLAND, Birger. Information Seeking and Subject Representation. **An Activity-theoretical approach to Information Science**. Westport & London: Greenwood Press, 2001.

KLEIBOER, M. Simulation methodology for crisis management support. **Journal of Contingencies and Crisis Management**, Nijmegen, v. 5, n. 4, p. 198-206, Dec. 1997.

LANCASTER, F. W. **Indexação e resumo**; teoria e prática. Brasília: Briquet de Lemos/Livros, 2004. 347p.

MANOLA, F., MILLER, E. **RDF primer**: W3C Recommendation, 2004. Disponível em: <http://www.w3.org/TR/rdf-primer/>. Acesso em: 05 mai. 2015.

MOURA. M.A. Informação, Ferramentas ontológicas e Redes Sociais: a interoperabilidade na construção de Tesauros Ontológicos**. IX Congresso ISKO** – Espanha; Valencia, marco, 2009.

MOURA, M.A. Folksonomias, Redes Sociais e a formação para o Tagging Literacy: desafios para a organização da informação em ambientes colaborativos virtuais**. Inf Inf**, Londrina, v. 14, n.1, p.25-4 5, 2009.

Mons, B; Velterop, J. Nanopublication in the e-science Era" **Workshop on Semantic Web Applications in Scientific Discourse** (SWASD 2009), 2009.

NANOPUB.ORG. Disponível em<http://nanopub.org/wordpress/>. Acesso em: 14 mar. 2014.

The Open PHACTS Nanopublication Guidelines. Disponivel em: <http://www.nanopub.org/guidelines/1.8/>. Acesso em: 14 mar. 2014.

VICENTE, P. O uso de simulação como metodologia de pesquisa em ciências sociais. **Cadernos EBAPE.BR**, v. 3, n. 1, mar. 2005.

WILSON, Patrick. **Two Kinds of Power: An Essay on Bibliographical Control**. Berkeley: University of California Press, 1968.

Aplicação de ontologias no processo de recuperação da informação aplicado em ambientes acadêmicos

Caio Saraiva Coneglian
Universidade Estadual Paulista
caio.coneglian@gmail.com

Elvis Fusco
Centro Universitário Euripides de Marília
fusco@univem.edu.br

José Eduardo Santarém Segundo
Universidade de São Paulo
santarem@usp.br

Introdução

A explosão de geração massiva de dados está testando a capacidade das mais avançadas tecnologias de armazenamento, tratamento, transformação e análise de informações. As áreas do tratamento e da recuperação da informação estão sendo desafiadas pelo volume, variedade e velocidade de uma inundação de dados semiestruturados e não estruturados de natureza complexa.

De forma que, que existe uma grande quantidade de dados espalhados pela Web, havendo uma grande dificuldade em organizar todas estas informações e assim, gerar conhecimento aos indivíduos. Bastos (2005) diz que a geração do conhecimento só

ocorre com a assimilação da informação organizada, pois é capaz de estabelecer relações de significados com os conhecimentos.

A Ciência da Informação, como diz Borko (1968), é uma disciplina que abrange a origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. Desta forma as áreas de recuperação da informação e organização do conhecimento se relacionam, de modo que, uma é capaz de complementar a outra em seus objetos de estudos.

De modo que, para tratar a questão de como organizar as informações e desta maneira apresentar resultados mais eficientes, são associados conceitos semânticos aos processo de Recuperação da Informação, com o propósito de adicionar significado aos conteúdos buscados em domínio específico, que permitem realizar a procura não mais por palavras chaves num processo de busca textual, mas sim por significado e valor, extraindo das páginas e serviços da Web informações de real relevância, descartando aquilo que é desnecessário. A partir disto, a ontologia aparece como solução na busca de inserir semântica neste processo.

Esta pesquisa tem como objetivo criar uma plataforma semântica de Recuperação de Informação na Web que permita a localização, armazenamento, tratamento, organização e recuperação de informações inseridos no contexto da Web nas mais variadas fontes informacionais na Internet que sirvam de base para uma arquitetura computacional que transforme a informação desagregada em um ambiente de conhecimento estratégico, relevante, preciso e utilizável para permitir aos usuários o acesso as informações com maior valor agregado, que consiga satisfazer as necessidades informacionais do usuário, aderindo uma semântica ao processo de Recuperação da Informação.

Metodologia

O presente estudo caracteriza-se como uma pesquisa descritiva e analítica, com base em análise documental, com característica exploratória, focalizando a implementação de uma plataforma de recuperação semântica da informação aplicada a Web. A construção do projeto foi dividido em três partes principais, levantamento bibliográfico, construção da ontologia e a implementação do protótipo.

Recuperação de informação

Os sistemas de informação tradicionais são incapazes de lidar de forma eficiente com todas as novas fontes de dados dinâmicas e de contextos múltiplos de informações que têm principalmente a Internet como plataforma.

São encontrados problemas em recuperar, padronizar, armazenar, processar e utilizar informações geradas por diversas fontes heterogêneas que servem de base para alimentar os sistemas de apoio à decisão das organizações.

Neste contexto faz-se o questionamento se os ambientes computacionais de informação realmente apresentam de forma completa as informações relevantes aos tomadores de decisão nas organizações.

Para isto foi proposta a criação de uma arquitetura de Recuperação de Informação no contexto da Web como pode ser visto na Figura 1.

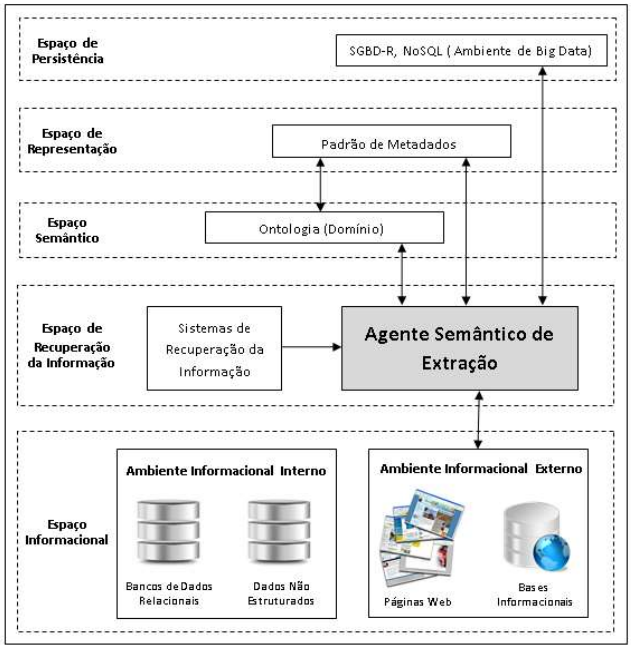


Figura 1: Contextualização da Arquitetura do Agente Semântico

Esta arquitetura foi proposta para que a recuperação da informação possa ser feita utilizando o espaço semântico. A arquitetura tem como centro o agente de recuperação de informação e integra-se com todos os outros elementos e camadas.

Esta arquitetura será usada em uma instituição de ensino superior, porque nesta área há muitos informação não está sendo usado várias vezes em inteligência competitiva.

Esta arquitetura distingue-se de outras, pois analisará todas as informações que serão recuperadas, verificando se as informações estão contidos dentro de um domínio, portanto, só serão extraídas informações relacionados ao problema que o usuário possui, de acordo com a ontologia definida.

Ontologia

A palavra ontologia vem de ontos (ser, ente) e logos (saber, doutrina), e de maneira estrita significa o “estudo do ser”. Surgiu do estudo de filósofos, ainda na época de Aristóteles, e era usada neste contexto para fazer uma abordagem do ser enquanto ser, ou seja do ser de uma maneira geral. Mais tarde ainda na filosofia, o termo ontologia passou a ser mais usado para saber aquilo que é fundamental ou irredutível, comum a todos os seres.

Ontologia é definida como uma teoria lógica que representa um vocabulário pretendido, ou seja, é uma contextualização de algo particular existente no mundo. Neste sentido observa-se que com uma ontologia você consegue definir contextos e domínios particulares do mundo (GUARINO, 1998)

Gruber (1995) diz que em um contexto de múltiplos agentes, a ontologia poderia definir o contexto, o vocabulário daquele domínio, servindo assim de base para a comunicação entre os agentes, e para conseguir fazer suas extrações no conhecimento em que eles estão presentes. Gruber (1995) ainda diz que a ontologia é uma especificação explícita de uma conceitualização.

Segundo Santarém Segundo (2010) a Ciência da Informação e a Ciência da Computação utilizou a ontologia quando se refere a aquisição de conhecimentos a partir de dados semiestruturados, utilizando da ontologia para aplicar técnicas e métodos, para processar as informações.

Santarém Segundo (2010) ainda diz que as ontologias vêm com o principal objetivo de ter um vocabulário compartilhado, onde essas informações possam ser

trocadas, e usadas para outros usuários. Sendo que estes usuários são tanto seres humanos quanto agentes inteligentes.

Para a construção da ontologia, é necessário utilizar uma metodologia que permita que a ontologia construída seja consistente. Para tal foi utilizado a metodologia proposta por Noy (2001a), onde é explicado que para a construção da ontologia devem ser seguidos sete passos.

Agente semântico de extração

A criação de um agente de software que agregue semanticamente as informações disponíveis na Web de um determinado domínio pode trazer para uma plataforma computacional subsídios para a criação de um ambiente informacional de apoio à decisão que dê uma visão mais ampla dos cenários internos e externos das informações de relevância na gestão organizacional.

Este agente deverá se comunicar com os espaços informacionais internos e externos da Web baseando suas buscas em regras ontológicas baseadas num padrão de metadados para realizar a extração semântica do domínio proposto e apoiará outros sistemas num contexto mais amplo de Recuperação da Informação.

A pesquisa utiliza o domínio de instituições de ensino superior como estudo de caso para aplicar a plataforma computacional proposta na arquitetura descrita na Figura 1. Como protótipo da ontologia, foi utilizado a disciplina de banco de dados, para buscar artigos científicos.

Notação Conceitual da Ontologia

O primeiro passo para a criação da ontologia, foi verificar quais são os termos da disciplina de banco de dados.

Após verificar quais eram estes termos, foi feito a estruturado a hierarquia entre as classes da ontologia, isto foi realizado segundo a experiência dos autores e de outros pesquisadores na área.

Foi elaborada a notação conceitual da ontologia, utilizando o software Protege (NOY, 2001b), como mostrado na Figura 3, sendo também construído o relacionamento entre as classes.

O agente irá utilizar esta ontologia para realizar uma tarefa, portanto esta ontologia que foi construída, é classificada como uma ontologia de tarefa, conforme a classificação de Mizoguchi (2003).

Uma ontologia de tarefa é uma ontologia que resolve um problema específico dentro de um domínio, ou seja, resolve o problema de pesquisa científica no domínio da disciplina de banco de dados.

Sendo após implementado a ontologia no software Protégé, criando o diagrama de classe com as suas propriedades, utilizando com linguagem o OWL (Web Ontology Language). Após a criação do arquivo do OWL, foi utilizado o software Owl2Java (ZIMMERMANN, 2009) transformado o OWL em classes Java; tornando desta forma a ontologia implementada.

Funcionamento do Agente Semântico

O agente de busca captura as informações por meio de páginas da web pré-definidos e usa a ontologia para classificar e após realizar uma busca de maneira semântica.

Figura 2: Funcionamento do Agente

A Figura 2 mostra o processo realizado pelo sistema. O usuário faz uma pesquisa sobre qualquer assunto relacionado a área de banco de dados, o agente extrai das bases de dados resumos relativos a este tema. Esses resumos deverão passar por um processo onde serão analisadas, levando em consideração as palavras contidas neste resumo que estão presentes no tema pesquisado pelo usuário. Desta forma é possível usar uma

ontologia construída, que trata de um tema específico na área de pesquisa científica, no caso na área da disciplina de banco de dados.

A sequência de tarefas realizadas por o agente é descrito como se seguir:

Extração da Informação

O agente extrai da página do IEEE Xplore (http://ieeexplore.ieee.org), os resumos, baseado na pesquisa que o usuário executa. Baseado na localização dos resumos no HTML na página, o agente extrai as informações, e transforma isto numa cadeia de String.

O processo do agente é divido em três fases: busca na página, extração dos títulos e resumos e devolução ao programa principal uma lista com os artigos.

- Busca na página HTML: esta primeira fase se caracteriza por realizar uma busca no sistema de busca do IEEE Xplore, e a extração do HTML que a página retorna como resultado desta busca.
- Extração de títulos e resumos da página: após o retorno do HTML, o agente extrai deste, o título e o resumo de cada artigo. Isto é possível por uma análise da página HTML, verificando as tags cujo os dados dos resumos e dos títulos estão inseridos.

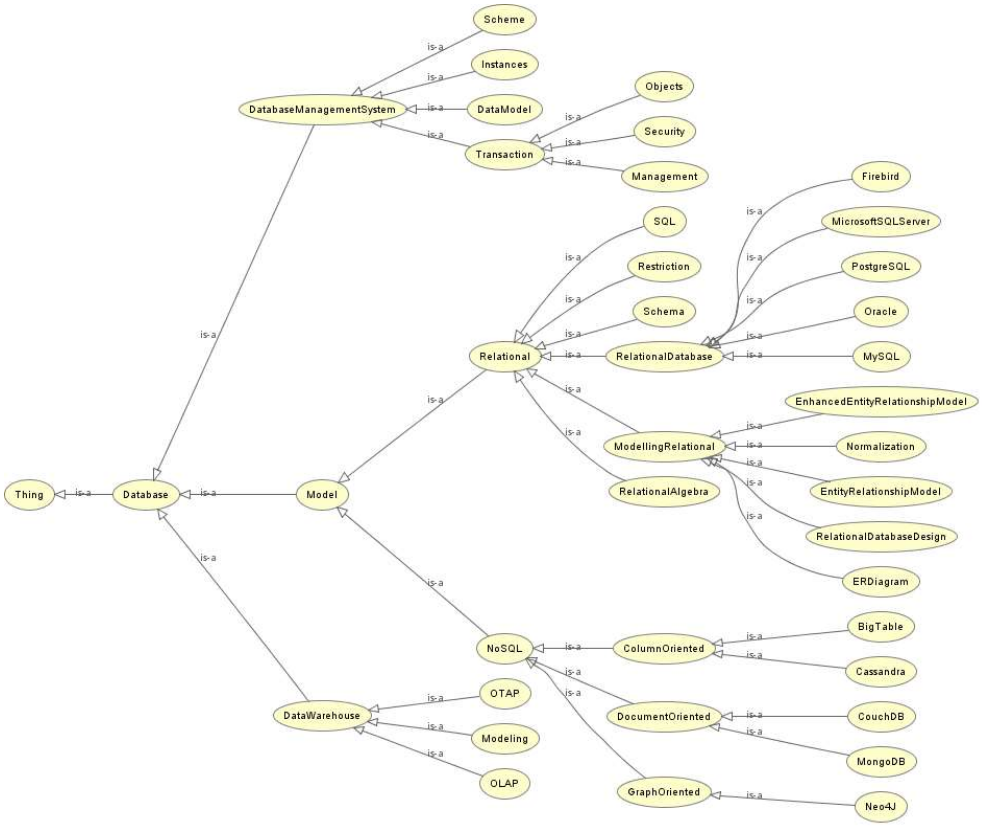


Figura 3: Relação de classes da ontologia

Desta forma, este robô de busca, consegue realizar uma extração sintática dos artigos contidos na base de dados do IEEE Xplore, pois, o robô de busca recupera os artigos que foram indexados pela própria base de dados, criando uma lista com todos os artigos que foram apresentados, para ser utilizado na ontologia.

Integração da Ontologia com o Agente de Extração

Para que o programa tenha de fato a semântica apresentada, o programa faz o uso da ontologia, para avaliar quais dos resultados que foram extraídos da base de dados, são de fato úteis, e tem relação com o contexto daquela busca.

Esta integração acontece em cinco momentos:

- Primeiramente, é verificado onde o termo pesquisado pelo usuário se encontra dentro da ontologia.
- Depois são obtidos, quais são as classes hierarquicamente superior e inferior ao termo pesquisado.
- Posteriormente é verificado dentro do resumo e do título dos artigos pesquisados, se contém ou não, os termos que fazem parte daquela hierarquia do termo pesquisado.

- Após, é realizado uma comparação entre quantidade de termos que estão na hierarquia e os que estão contidos dentro do resumo e do título daquele artigo. Resultando assim uma porcentagem da quantidade de termos que estão na hierarquia, que estão dentro do resumo e do título daquele artigo.
- Por fim, é apresentado ao usuário todos os artigos que alcançaram uma porcentagem acima dos 35%.

Resultados

Como teste para averiguar se o sistema está extraindo e verificando a semântica dos artigos extraídos, foi feita uma busca com o usuário pesquisando pelo termo “Datawarehouse”.

A hierarquia do termo Datawarehouse são os termos: Database, Datawarehouse, OLAP, OTAP e modeling.

No caso de 25 artigos, 7 foram os que atenderam aos requisitos, sendo estes apresentados aos usuários. Para visualizar como o programa faz a análise dos resumos e dos títulos, abaixo na figura 4, é apresentado um artigo dos que atenderam aos requisitos.

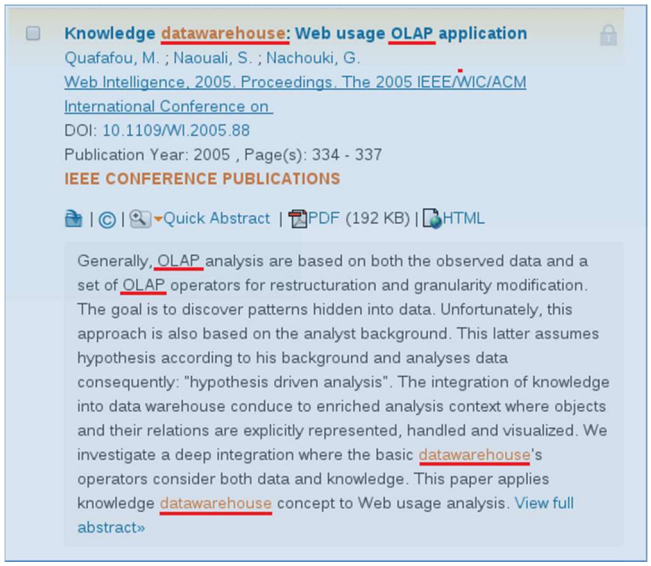

Knowledge datawarehouse: Web usage OLAP application
Quafafou, M. ; Naouali, S. ; Nachouki, G.
Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on
DOI: 10.1109/WI.2005.88
Publication Year: 2005 , Page(s): 334 - 337
IEEE CONFERENCE PUBLICATIONS
Quick Abstract | PDF (192 KB) | HTML
Generally, OLAP analysis are based on both the observed data and a set of OLAP operators for restructuration and granularity modification. The goal is to discover patterns hidden into data. Unfortunately, this approach is also based on the analyst background. This latter assumes hypothesis according to his background and analyses data consequently: "hypothesis driven analysis". The integration of knowledge into data warehouse conduce to enriched analysis context where objects and their relations are explicitly represented, handled and visualized. We investigate a deep integration where the basic datawarehouse's operators consider both data and knowledge. This paper applies knowledge datawarehouse concept to Web usage analysis. View full abstract»

Figura 4: Exemplo de um artigo analisado

Este artigo apresentou 2 dos 5 termos da hierarquia da ontologia. Neste caso apresentou os termos OLAP e Datawarehouse. Na figura 4 está grifado em vermelho os termos que foram encontrados pelo programa.

Considerações finais

Este trabalho apresenta o uso de ontologias na melhoria do processo de Recuperação de Informação.

O objetivo desta pesquisa é aderir semântica ao processo de Recuperação da Informação, utilizando das informações dentro do contexto do Big Data, para realizar um processo que agregue mais valor às buscas realizadas pelo usuário.

Para comprovar este objetivo, foi utilizado o domínio de pesquisas científicas, em que o usuário ao realizar uma busca em bases de dados de artigos científicos, se depara com o problema de ter uma quantidade muito grande de documentos, sendo que boa parte destes, não atendem às necessidades que o usuário possui.

Foi, então, criado uma ontologia e um robô de buscas e realizada a conexão entre estes para alcançar desta maneira o objetivo inicial.

Após a realização de testes, foi observado que o uso de ontologia para o agente de pesquisa é uma maneira eficaz para se obter informações de valor e conseguir atender as necessidades informacionais do usuário.

A ontologia pode ser eficiente no presente processo, porque se torna uma forma de organizar a informação semântica, e assim, apenas a informação significativa será apresentada ao usuário.

Embora o termo Web Semântica é usado já a alguns anos, ainda existe uma limitação em seu uso, porque grande parte da Web está organizada de uma forma sintática, em que a maioria das páginas são criadas para que apenas o ser humano consiga ler o que lá está escrito, sem serem estruturadas de uma maneira que agentes computacionais consigam extrair os dados ali contidos dentro de um contexto, com um significado implícito dentro do HTML.

O agente de extração consegue retirar os documentos da Web e um programa consegue por meio do uso de ontologia, tratar as informações, conseguindo assim apresentar resultados mais relevantes aquele usuário.

Desta maneira os resultados obtidos com a utilização do protótipo desenvolvido, consegue refinar bastante a quantidade de artigos apresentados aos usuários. Esta pesquisa, busca portanto, fazer com o que o usuário obtenha, em um processo de Recuperação de Informação, resultados mais expressivos e que apresente maior valor. Assim, o usuário conseguirá avaliar informações mais

expressivas, e não perderá tempo com aqueles dados que não tem atende suas necessidades.

Portanto, para tratar a questão de como inserir uma inteligência na recuperação de páginas Web que não apresentam uma contextualização de suas informações, esta pesquisa propõe que o processo de aderir semântica a estas páginas ocorra fora da Web, ou seja, a extração das páginas ocorra de maneira sintática, e a partir do que foi extraído, ocorra uma análise das informações, inserindo desta forma semântica a este processo. Este método se mostrou muito eficiente, pois consegue de fato realizar uma busca mais inteligente, que vai além de simples fórmulas de buscas, que observam apenas a sintaxe dos textos, e consegue analisar o contexto na qual os documentos extraídos estão inseridos, e assim visualizar se aquele documento atende ao que o usuário necessita.

Agradecimentos

O trabalho apresentado no documento foi financiado pela FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), processo 2013 / 16369-3.

Referências

BASTOS, Flavia Maria. **Organização do conhecimento em bibliotecas digitais de teses e dissertações: análise da aplicabilidade das teorias macroestruturais para categorização de áreas de assunto.** 2005.

BORKO, Harold. Information science: what is it? **American documentation**, v. 19, n. 1, p. 3-5, 1968.

BORST, Willem Nico. **Construction of engineering ontologies for knowledge sharing and reuse**. Universiteit Twente, 1997.

GRUBER, Thomas R. **Toward principles for the design of ontologies used for knowledge sharing?.** International journal of human-computer studies, v. 43, n. 5, p. 907-928, 1995.

GUARINO, Nicola (Ed.). **Formal ontology in information systems: Proceedings of the first international conference (FOIS'98), June 6-8, Trento, Italy**. IOS press, 1998.

MIZOGUCHI, Riichiro. Tutorial on ontological engineering. **NEW GENERATION COMPUTING-TOKYO-**, v. 21, n. 4, p. 363-364, 2003.

NOY, Natalya F. et al. **Ontology development 101: A guide to creating your first ontology.** 2001a.

NOY, Natalya F. et al. **Creating semantic web contents with protege-2000**.IEEE intelligent systems, v. 16, n. 2, p. 60-71, 2001b.

SANTAREM SEGUNDO, José Eduardo. **Representação Iterativa. Um Modelo para Repositórios Digitais**. Marília, São Paulo, Brasil, 2010.

ZIMMERMANN, M. **Owl2Java-A Java Code Generator for OWL**, 2009.

Atuação bibliotecária na análise de assunto de teses nas dimensões teóricas da catalogação de assunto e indexação

Roberta Cristina Dal'Evedove Tartarotti
Universidade Estadual Paulista
roberta_tartarotti@yahoo.com.br

Paula Regina Dal'Evedove
Universidade Federal de Pernambuco
paula.dalevedove@ufpe.br

Mariângela Spotti Lopes Fujita
Universidade Estadual Paulista
fujita@marilia.unesp.br

Introdução

No âmbito da Organização e Representação do Conhecimento (ORC), destaca-se o Tratamento Temático da Informação (TTI) como uma subárea fundamental ao lidar com as questões acerca da "análise, descrição e representação do conteúdo dos documentos, bem como suas inevitáveis interfaces com as teorias e sistemas de armazenamento e recuperação da informação" (BARITÉ, 1997, p. 124).

Guimarães (2009, p. 106) esclarece que em busca de uma fundamentação teórica e estabelecimento como campo de pesquisa, historicamente o TTI apresenta-se sob três correntes teóricas: a análise documental *(analyse documentaire*, de concepção francesa); a catalogação de assunto (*subject cataloguing,* de concepção norte-americana) e a indexação (*indexing*, de concepção inglesa).

Enquanto a catalogação de assunto é a dimensão do processo de catalogação responsável pela análise de assunto de itens informacionais, bem como o fornecimento de notações de classificação e cabeçalhos de assunto que representem o assunto dos itens – voltada para os catálogos coletivos *online*; a indexação é um processo formado por subprocessos/etapas que tem como objetivo identificar o conteúdo de um documento, por meio de uma metalinguagem construída (linguagem de indexação) com o intuito de promover a efetiva recuperação da informação – voltada para as bases de dados de áreas científicas especializadas.

Conforme Fujita, Rubi e Boccato (2009), enquanto a catalogação de assunto está essencialmente ligada à construção de catálogos de bibliotecas, a indexação está ligada à construção de índices de bibliografias em serviços de informação bibliográficos que produzem bases de dados, pois

> [...] na catalogação do livro, o seu conteúdo é tratado no todo, e os assuntos são fornecidos em uma escala limitada (um número de classificação para arranjo nas estantes e um ou dois cabeçalhos de assunto para acesso por meio do catálogo). Já na indexação de outros materiais, a tendência é o detalhamento, em que há maior generosidade no fornecimento de termos para o acesso por assunto (FUJITA; RUBI; BOCCATO, 2009, p. 31).

Embora processos de TTI distintos, a qualidade tanto da catalogação de assunto quanto da indexação está relacionada à capacidade de reconstruir o assunto tratado em um documento em conceitos para recuperação posterior pelo usuário do sistema de informação. O objetivo da prática profissional tanto do indexador como do catalogador de assunto é tornar conhecido o assunto de um documento, ou seja, representar seu conteúdo por meio de conceitos significativos.

Como responsáveis pelo armazenamento, localização, busca e seleção, sob demanda, de dados pertinentes a um determinado assunto, os catálogos coletivos *online* e as bases de dados de bibliotecas universitárias são considerados sistemas de recuperação da informação, em que a qualidade da representação temática está diretamente relacionada à qualidade da análise temática, considerando sua complexidade inerente ao fazer profissional (TARTAROTTI, 2014, p. 23).

Desse modo, o objetivo da pesquisa foi investigar a atuação bibliotecária em sistemas de recuperação da informação na área da Saúde, para observação dos processos cognitivos de análise do documento do tipo tese na catalogação de assunto

no catálogo coletivo *online* DEDALUS (USP) e do processo de indexação na base de dados LILACS (BIREME). Por meio dos aportes teórico-metodológicos, foi possível observar se o bibliotecário do TTI efetua a análise de assunto no processo de catalogação de assunto em catálogos coletivos, diferentemente da forma que realiza a indexação em bases de dados.

Procedimentos metodológicos

Para observação da atuação bibliotecária, como metodologia qualitativa para coleta de dados foi utilizada a técnica do Protocolo Verbal (PV) na modalidade Protocolo Verbal Individual (PVI). A técnica, cujos precursores são Ericsson e Simon (1993), consiste em um método de análise introspectivo cujo objetivo é a gravação da exteriorização verbal do pensamento durante a realização de uma determinada atividade de leitura. O "pensar alto" do sujeito é gravado e transcrito literalmente, produzindo protocolos verbais.

O universo da pesquisa contemplou três Bibliotecas universitárias públicas da Universidade de São Paulo: Biblioteca da Faculdade de Odontologia (FO), Biblioteca da Faculdade de Saúde Pública (FSP) e Biblioteca do Instituto de Psicologia (IP). A amostra de três bibliotecas universitárias foi definida após análise dos sistemas de recuperação da informação e contato com as Bibliotecas participantes, verificando se as mesmas realizavam tanto a catalogação de assunto de teses no catálogo coletivo *online* DEDALUS quanto à indexação na base LILACS, para que a análise pudesse proporcionar a comparação dos procedimentos e dificuldades em torno destas atividades.

No caso da Biblioteca da Faculdade de Saúde Pública (FSP), o mesmo profissional realizou tanto a catalogação de assunto no catálogo coletivo *online* DEDALUS quanto a indexação de teses na base de dados LILACS, tendo sido realizado, dessa maneira, dois protocolos com o mesmo bibliotecário.

No tocante à infraestrutura material para a realização da aplicação do PVI, as Bibliotecas atenderam os quesitos necessários. Para tanto, foi utilizado o próprio ambiente de trabalho do bibliotecário catalogador e indexador, para que assim a coleta de dados fosse capaz de refletir a realidade das atividades de catalogação e indexação sem interferências.

As coletas de dados foram realizadas uma única vez em cada biblioteca universitária, o que deu origem a seis protocolos verbais, com cinco sujeitos participantes. A técnica do PVI possibilitou a comparação da prática profissional da catalogação de assunto em catálogo coletivo *online* e do processo de indexação em base de dados científica especializada.

A seguir, são apresentados, sinteticamente, os procedimentos anteriores, durante e posteriores à aplicação do PVI:

Procedimentos anteriores à aplicação do PVI

a) Definição do universo da pesquisa;
b) Seleção dos sujeitos;
c) Definição das tarefas;
d) Seleção do texto-base;
e) Conversa informal com os sujeitos participantes;
f) Familiarização dos sujeitos com a técnica do PVI.

Procedimentos durante a aplicação do PVI para catalogação de assunto/indexação

a) Gravação do “Pensar Alto” dos sujeitos participantes.

Procedimentos após a aplicação do PVI

a) Transcrição literal das gravações das falas dos sujeitos participantes;
b) Leitura detalhada dos dados em busca de fenômenos significativos e recorrentes para elaboração de categorias de análise;
c) Elaboração das categorias de análise;
d) Retorno aos dados para retirar trechos da discussão que exemplifiquem cada categoria de análise.

Resultados

A análise dos resultados pautou-se em duas *categorias de análise*: Análise de assunto/Identificação de conceitos e Representação/Tradução de conceitos. Em relação à categoria **Análise de assunto/Identificação de conceitos,** foi possível

observar que todos os profissionais, sejam catalogadores de assunto ou indexadores, realizaram a análise do conteúdo do documento, entretanto, em diferentes níveis de profundidade. Pautaram-se ainda em determinadas partes do documento (estrutura textual), em busca da identificação e seleção de conceitos e, por vezes, em busca da compreensão do conteúdo do mesmo.

QUADRO 1: Partes do texto utilizadas pelos profissionais durante a análise de assunto/identificação de conceitos de tese.

CAT - FO	IND - FO	CAT - FSP	IND - FSP	CAT - IP	IND - IP
Título	Título	Título			Título
Resumo	Resumo	Resumo	Resumo	Resumo	Resumo
	Palavras-chave do resumo				
Palavras-chave da ficha catalográfica				Palavras-chave da ficha catalográfica	
Introdução					
Figuras					
			Sumário		

Fonte: elaborado pelas autoras.

Em busca de conceitos significativos que pudessem representar o conteúdo do documento do tipo tese, verificou-se que o catalogador de assunto da FO utilizou as seguintes partes do texto: título, resumo, palavras-chave da ficha catalográfica, introdução e figuras, ressaltando durante o PVI a importância da leitura técnica, enquanto que o indexador optou pelo título, o resumo e as palavras-chave do resumo. Além disso, o indexador explicou que na biblioteca em que atua existem estagiários da área de Odontologia que auxiliam em momentos de dúvida dos profissionais quanto aos termos utilizados.

Na FSP, o catalogador de assunto utilizou o título e o resumo. O mesmo profissional, que também realizou o processo de indexação, utilizou o resumo e o sumário da tese. Visto que as duas atividades foram realizadas na sequência, é possível que, durante a indexação, o profissional tenha considerado desnecessário analisar novamente o título da tese. Já no IP, o catalogador de assunto utilizou

rapidamente o resumo para confirmar as palavras-chave já atribuídas na ficha catalográfica da tese, enquanto que o indexador pautou-se no título e no resumo para identificar os conceitos relevantes que representassem o conteúdo do documento.

No que se refere às diretrizes profissionais para a análise de assunto (Norma ISO 5963 para indexação e Classificação Decimal de Dewey para classificação), Mai (1997, p. 60) alerta que não declaram nada sobre como o exame do conteúdo temático dos documentos, indicando apenas as fontes potenciais para localizar o assunto. Além disso, nenhuma destas orientações comuns e tão utilizadas na atuação profissional esclarece como o indexador deve determinar o conteúdo de assunto com base na análise das necessidades dos usuários e dos contextos sóciocognitivo e sociocultural.

De acordo com a Norma 12.676 (ABNT, 1992), uma tradução da Norma ISO 5963, as partes do texto que devem ter uma maior relevância são: título e subtítulo; resumo se houver; sumário; introdução; ilustrações, diagramas, tabelas e seus títulos explicativos; palavras ou grupos de palavras em destaque (sublinhadas, impressas em tipo diferente, etc.); e referências bibliográficas. Entretanto, destaca-se a falha que a referida norma possui por não identificar quais são os tipos de conceitos que podem ser encontrados em cada parte textual indicada. De acordo com Fujita (2003), além das indicações que a Norma faz, é possível que o profissional recorra também às primeiras frases dos capítulos para identificar os conceitos representativos dos documentos.

Todos os profissionais nas bibliotecas pesquisadas utilizaram o resumo como principal parte do texto para identificar e selecionar os conceitos representativos do documento do tipo tese. Desse modo, apreende-se que o resumo é parte essencial neste processo, antecedendo até mesmo o título em nível de importância. Um resumo bem escrito deve ressaltar os seguintes pontos de forma sintética: contextualização da pesquisa, justificativa, problemática, objetivo geral e objetivo(s) específico(s), resultados e considerações finais. Outras partes da estrutura textual de teses, como título, palavras-chave atribuídas pelo autor, sumário, introdução, ilustrações (figuras, gráficos e tabelas) como também a conclusão são elementos relevantes para a identificação e seleção de conceitos.

No tocante à categoria **Representação/Tradução de conceitos**, o catalogador da FO explicitou que o *Sistema de Sugestões do Vocabulário Controlado do SIBi/USP* (SIBIX) não alcança o nível de especificidade abordado na tese. Desse modo, termos

que poderiam ser atribuídos ao documento não puderam ser utilizados devido à restrição da linguagem documental utilizada na etapa de representação e tradução de conceitos. Entretanto, a mesma situação ocorreu com o indexador da mesma instituição, pois ao tentar traduzir para a linguagem de indexação termos identificados e selecionados durante a análise de assunto, no DeCS os termos não foram localizados.

O catalogador da FSP denotou que não seria utilizada, na etapa de tradução, a mesma quantidade de termos identificados e selecionados. O catalogador do IP, ao analisar a ficha catalográfica da tese, expôs que, pela sua experiência profissional, foi possível verificar se os termos ali contidos fazem ou não parte do vocabulário controlado, ou seja, do SIBIX.

O indexador do IP, ao mesmo tempo em que identificou e selecionou os conceitos significativos do conteúdo da tese, também realizou a pesquisa dos descritores na linguagem de indexação específica da área de Psicologia. A linguagem de indexação também auxiliou o profissional na compreensão do conteúdo documental.

Considerações finais

A leitura documental é considerada a fase mais importante do processo de análise de documentos, sendo o sucesso nesta fase inicial que garantirá a qualidade do restante do processo. É um fato comprovado a dificuldade apresentada pelo indexador durante a leitura documental em relação à “identificação e seleção de conceitos representativos do assunto do documento” (SILVA; FUJITA, 2004).

> [...] a representação por conceitos na análise de assunto para identificação e seleção de conceitos é diferente da representação realizada na tradução porque, em primeiro lugar, é realizada durante a análise de conteúdo do documento, ou seja, com o documento na íntegra e em segundo lugar, porque utiliza conceitos com os quais são representados os termos e assim identificada a tematicidade intrínseca do documento. Na tradução, de outro modo, a representação se dá com os termos extraídos do documento analisado, portanto, fora do contexto documentário em que o autor desenvolveu o conteúdo significativo e utiliza uma linguagem *documental* que representa o vocabulário de áreas de especialidade com a finalidade de compatibilizar os termos identificados com os termos buscados pelo usuário (FUJITA, 2013, p. 50, alteração nossa).

No que diz respeito ao **processo**, entende-se que durante a catalogação de assunto/indexação, as etapas não precisam, necessariamente, ser realizadas de maneira sequencial, visto que o indexador profissional já familiarizado com o processo pode realizá-las simultaneamente (FUJITA; RUBI; BOCCATO, 2009, p.24). Do mesmo modo, quanto mais familiarizado com a área científica especializada no qual atua, mais estas etapas são sobrepostas durante o processo de catalogação de assunto pelo profissional no âmbito de bibliotecas universitárias.

De acordo com Mai (2000, p. 280), a ação tomada por indexadores com diferentes níveis de experiência pode não ser exatamente a mesma, mas as etapas e elementos do processo de indexação de assunto são considerados fundamentais para qualquer processo de indexação. Nesta investigação verificou-se, ainda, que o profissional realizou a catalogação de assunto e a indexação de forma distinta, visto que nas bases de dados o processo de indexação foi realizado em um maior nível de especificidade do que a catalogação de assunto em catálogo coletivo *online*.

Quanto ao **catalogador de assunto/indexador**, tanto o catálogo coletivo *online* como as bases de dados de áreas científicas especializadas permitem a visibilidade do conteúdo informacional das bibliotecas universitárias. No entanto, a determinação do(s) assunto(s) nestes sistemas de recuperação da informação de bibliotecas universitárias envolve a subjetividade e bom senso profissional, considerando-se que tanto o catalogador de assunto como o indexador geralmente não são especialistas na área científica especializada em que atuam.

Já em relação ao **contexto** das bibliotecas pesquisadas, apreende-se a necessidade de uma política de indexação formalizada em um manual de indexação que norteie tanto a catalogação de assunto no catálogo coletivo *online*, assim como a indexação em bases de dados científicas especializadas, cujos registros são inseridos pelos profissionais que neles atuam. Diretrizes e procedimentos de catalogação de assunto/indexação bem estabelecidos e seguidos pelos profissionais podem contribuir para a melhoria da qualidade da informação recuperada.

Por fim, ao lançarmos o olhar para o profissional que realiza o tratamento temático da informação em bibliotecas universitárias, contribuímos para que as incoerências ou omissões nos processos de catalogação de assunto e indexação sejam amenizadas, visando maior consistência na criação de produtos informacionais.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676*:*** Métodos para análise de documentos - determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992. 4 p.

BARITÉ, M. Referenciales teóricos vigentes en el área de tratamiento temático de la información y su expresión metodológica. Porto Alegre: **ABEBD**, 1998. 7 p. [Relatório técnico do II Encontro de Dirigentes dos cursos superiores de Biblioteconomia dos países do Mercosul, Buenos Aires, nov. 1997].

ERICSSON, K. A.; SIMON, H. A. **Protocol analysis*:*** verbal reports as data. Rev. ed. Cambridge: MIT Press, 1993. 496 p.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação.** Campinas, v.1, n.1, p.60-90, jul./dez. 2003.

FUJITA, M. S. L.; RUBI, M. P.; BOCCATO, V. R. C. As diferentes perspectivas teóricas e metodológicas sobre indexação e catalogação de assuntos. In: FUJITA, M. S. L. (Org.). **A indexação de livros:** a percepção de catalogadores e usuários de bibliotecas universitárias. São Paulo: Cultura Acadêmica, 2009. p. 19-42.

GUIMARAES, J. A. C. Abordagens teóricas de tratamento temático da informação (TTI): catalogação de assunto, indexação e análise documental. **Ibersid***,* Zaragoza, p. 105-117, 2009.

ISO. **Documentation: methods for examining documents, determining their subjects, and selecting indexing terms.** Suíça: ISO, 1985. 5 p.

MAI, J-E. The concept of subject: on problems in indexing, In: **Knowledge Organization for Information Retrieval.** Proceedings of the 6th International Study Conference on Classification Research, v. 6, p. 60-67, 1997.

MAI, J-E. Deconstructing the indexing process. **Advances in Librarianship***,* v. 23, p. 269-298, 2000.

TARTAROTTI, R. C. D. E. **Atuação bibliotecária no tratamento temático da informação em unidades informacionais:** um estudo comparativo qualitativo-quantitativo. 2014. 277 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos - UFSCar, São Carlos, 2014.

A contribuição da identificação documental para a organização do conhecimento em arquivos pessoais

Gabrieli Aparecida da Fonseca
Universidade Estadual Paulista
E-mail: gabrieli.arq@gmail.com

Sonia Maria Troitiño Rodriguez
Universidade Estadual Paulista
Email: smtr@marilia.unesp.br

1 Introdução

A identificação documental, como etapa preliminar à classificação, é um procedimento arquivístico de suma importância para a organização de fundos documentais e, consequentemente, para a organização do conhecimento deles emanado. Contudo, o que pese este estudo versar sobre a gênese documental, em busca do lastro informacional proporcionado pelo princípio da proveniência, relaciona-se também com campos de estudos análogos, como a tipologia documental. Desse modo, configura-se como um estudo que busca articular informações advindas do entendimento do contexto de produção e acumulação dos documentos aliado à análise tipológica, em busca dos sentidos assumidos pela informação registrada e ratificados pela organicidade em seu meio.

Considerando que para explicitar a origem dos documentos é necessário ir além do estudo da história administrativa da instituição e abordar igualmente o sistema de produção documental através do estudo das séries dele decorrente, a identificação documental pode vir a contribui em muito para a efetiva aplicação do princípio de proveniência.

Segundo Ana Célia Rodrigues, a identificação documental

> […] trata-se de uma tarefa preliminar e necessária às funções da classificação, avaliação, descrição, planejamento da produção

> documental. [...] Neste sentido, é um trabalho de pesquisa e de crítica sobre a gênese documental (RODRIGUES, 2011, p. 118-119).

Contudo, ainda é uma área de estudo pouco difundida, de modo que a maior parte das pesquisas desenvolvidas sobre a temática, em geral, partem de estudos de caso sobre fundos de caráter público. Como consequência, as diversas pesquisas sobre identificação documental, nem sempre apresentam posições consensuais dentro do campo da ciência da informação – sendo os estudos dela derivados majoritariamente desenvolvidos por pesquisadores espanhóis, como Concepción Mendo Carmona ou Maria Luísa Conde Villaverde, entre outros.

Por outro lado, na Arquivologia existe diversos trabalhos investigativos cuja temática recai sobre a questão dos arquivos pessoais. Entretanto, a pesquisa aqui apresentada se propõe a uma abordagem pouco usual, tendo por tema os arquivos pessoais, visando entender como a identificação documental pode contribuir para a organização da informação de acervos privados, inclusive ao problematizar a origem e formas de acumulação que fazem com que tais arquivos e seus respectivos documentos sejam pessoais. Dessa forma, incorporamos ao cerne da discussão o conceito *Total Archives*, defendido pelo autor Terry Cook (1998), o qual discute os benefícios da integração informacional entre arquivos públicos e pessoais.

Diante desses pressupostos, buscamos debater o modo como a identificação documental pode contribuir para evidenciar as fronteiras que separam arquivos privados em pessoais e institucionais. Assim, o estudo de caso sobre os procedimentos de organização da informação desenvolvidos no Instituto Fernando Henrique Cardoso-iFHC vem a complementar nossa pesquisa ao analisar o tratamento documental sobre Fundo Fernando Henrique Cardoso, em busca da fundamentação proporcionada pela práxis em arquivos aliada às bases teóricas de nosso campo de estudo.

O Instituto Fernando Henrique Cardoso foi criado em 2004, pelo ex-presidente da república Fernando Henrique Cardoso, a fim de reunir toda sua documentação de cunho particular, assim como a de sua esposa, a socióloga Ruth Cardoso. Os documentos sob custódia do instituto são de suma importância, pois apesar de sem de origem pessoal, representam parte da história e memória do país. Entre os diversos tipos documentais encontrados no acervo estão cartas, anotações, livros, textos, fotografias, notas preparatórias de viagens, pareces técnicos, relatórios, agendas, certificados, etc.

A eleição desse acervo para a análise aqui posta se deve a proposta metodológica de organização arquivística eleita pelo instituto. Por ser a identificação documental um campo de estudo ainda muito recente e pouco explorado, são raras as instituições que a incorporam conceitualmente em sua metodologia de trabalho. Contudo, o iFHC pode ser considerado destaque devido ao trabalho de organização realizado com a aplicação da identificação documental. O pioneirismo do método empregado foi relatado no livro *Tempo e Circunstância: a abordagem conceitual dos arquivos pessoais*, pelas autoras Ana Maria de Almeida Camargo e Silvana Goulart (2007). O livro traz uma reflexão acerca dos arquivos pessoais, a partir da exposição do trabalho de organização realizado pelo iFHC, especialmente no que se refere à identificação e descrição documental.

Os conceitos e delimitações que diferenciam arquivos públicos de privados, sejam institucionais ou pessoais, são bastante polêmicos. Divergências essas intensificadas quando se trata de acervos de políticos, pois estes costumam ser constituídos tanto por documentos de foro íntimo, quanto por documentos criados durante o cargo público exercido. Nesse sentido, a identificação documental pode colaborar para evidenciar a natureza privada ou não do documento analisado.

2 Total archives e identificação documental

O *Total Archives*, teoria canadense que visa integrar arquivos privados a arquivos públicos de forma a complementá-los, surge diante da compreensão de que os diferentes contextos de produção documental são igualmente representativos da sociedade da qual decorrem. Segundo Terry Cook, um dos principais estudiosos do *Total Archives*, há duas semelhanças entre Arquivos Pessoais e Arquivos Públicos:

> O documento-proposta do Seminário Internacional sobre Arquivos Pessoais reconhece duas similaridades importantes entre os arquivos pessoais e os públicos. Primeiro, ambos são artefatos de registro derivados de uma atividade; os arquivos são evidências das transações da vida humana, seja ela organizacional, e por conseguinte oficial, seja individual, e portanto pessoal. Diversamente de livros, programas de televisão ou obras de arte, eles não são intencionalmente criados por motivos próprios, com a possível exceção dos textos autobiográficos, mas surgem, antes, dentro de um contexto, como parte de alguma outra atividade ou necessidade, seja pessoal, seja institucional. Em segundo lugar, os arquivistas, tanto nos arquivos públicos quanto nos pessoais, frequentemente usam procedimentos técnicos e métodos práticos semelhantes, em termos de como acessam, descrevem, armazenam fisicamente e conservam

> os arquivos e os colocam à disposição para fins de pesquisa (COOK, 1998, p. 131).

Nessa perspectiva, não há distanciamento, mas antes aproximação e intercomunicação entre registros oficiais e os produzidos paralelamente à eles, em âmbito privado. Desse modo, fica subentendido que arquivos pessoais servem de complemento aos institucionais públicos e vice-versa.

De acordo com o que sugere o *Total Archives*, os arquivos pessoais representam o preenchimento de lacunas informacionais nos arquivos públicos e institucionais, pois traz à tona a documentação que ficou oculta por se referir em primeiro plano as particularidades do indivíduo, mas que quando articuladas a outros registros recuperam relações subliminares e dão novo sentido aos fatos registrados de modo pontual. Isso ocorre pelo fato do indivíduo, enquanto ser da sociedade, interagir simultaneamente, ao longo de sua vida, tanto em âmbito público quanto privado, sendo os documentos decorrentes dessa interação representativos de sua própria atuação.

Ao refletir sobre o papel da identificação documental no processo de organização do conhecimento, alinhada à perspectiva fornecida pelo *Total Archives*, pudemos constatar que a identificação documental é crucial para realização de uma organização satisfatória de arquivos pessoais, tal qual a de arquivos institucionais: 1º) por sua capacidade de evidenciar a natureza e origem dos documentos, conforme a proveniência; 2º) por auxiliar na construção do plano de classificação ao esclarecer as relações orgânicas entre os diferentes níveis hierárquicos; e 3º) por contribuir para o estabelecimento de séries documentais tipológicas, através do reconhecimento de espécies e funções presentes nos registros das ações humanas no âmbito da vida privada.

Enfim, compartilhamos a posição de Fonseca (2015) quando afirma que a identificação é um meio de potencializar o tratamento documental, já que antecipa questões necessárias ao trabalho posterior a ela.

3 Considerações finais

Ao refletir sobre os valores informacionais contidos nos documentos pessoais do intelectual e ex-presidente da república Fernando Henrique Cardoso, foi possível constatar que a Identificação Documental pode ser utilizada na organização de

arquivos pessoais de modo a explicitar as especificidades desses acervos, através da definição dos tipos documentais, proveniência e acumulação dos documentos. No caso da organização analisada, o iFHC, a aplicação da identificação documental no tratamento documental demonstrou aspectos positivos, especialmente pela contribuição fornecida às demais funções arquivísticas, como classificação e descrição, tanto do ponto de vista operacional, quanto semântico.

Ainda hoje, os conceitos e as delimitações que diferenciam arquivos públicos de privados, sejam institucionais ou pessoais, são bastante polêmicos, uma vez que as características peculiares a cada um deles confundem-se. Dessa forma, pesquisas sobre arquivos pessoais e institucionais são necessárias a fim de pontuar como a identificação documental pode contribuir para a melhor organização e valorização desse tipo de documentação, através do estudo da gênese e tipologia documental.

Ana Maria de Almeida Camargo (2009) evidencia o cuidado necessário ao se trabalhar com arquivos pessoais. De acordo com a autora,

> Se a utilização de rótulos universais para a caracterização desses arquivos prepara perigosas armadilhas para os profissionais que deles se servem – colocando num mesmo plano, espécies, formas, gêneros, assuntos e formatos –, tem ainda mais dois efeitos perversos: compromete sua organicidade e sinaliza a renúncia ao caráter probatório que sua funcionalidade originária lhes proporciona (CAMARGO, 2009, p. 31).

Fica assim exposta a importância da identificação, uma vez que proporciona o entendimento sobre os documentos para além de suas formas, gêneros, assuntos e formatos, conforme citado, indo ao encontro de sua produção e acumulação. Cumpre assim, seu papel como instrumento para organização do conhecimento.

No entanto, o trabalho de identificação pode ser muito mais complexo em arquivos pessoais. Apesar das similaridades existentes entre arquivos de origem privada e pessoal ser grandes, este último se caracteriza pelo fato de apresentar grande número de tipos documentais não diplomáticos de foro íntimo, como cartas, cartões-postais, diários pessoais, entre outros, em oposição aos institucionais.

As especificidades que tornam a organização dos arquivos pessoais mais complexas, não devem dificultar ou impedir que os mesmos sejam tratados tal como arquivos que são – conforme os preceitos da arquivologia.

Dessa forma, a identificação documental nos arquivos pessoais, embora complexa devido a inexistência de um consenso na área, se faz útil e necessária, pois

através dela torna-se possível evidenciar a proveniência e organicidade, facilitando as etapas posteriores de organização da informação.

Referências

BRASIL, Congresso Nacional. **Lei nº 8.394, de 30 de dezembro de 1991**. Diário Oficial da União, de 6 de janeiro de 1992. Disponível em: <http://www.conarq.arquivonacional.gov.br/cgi/cgilua.exe/sys/start.htm?from%5Finfo%5Findex=11&infoid=97&sid=52> Acesso em : 16 de dez. de 2013.

CAMARGO, Ana Maria de Almeida. Arquivos pessoais são arquivos. **Revista do Arquivo público Mineiro**, p. 27-39. Disponível em: <http://www.siaapm.cultura.mg.gov.br/acervo/rapm_pdf/2009-2-A02.pdf>Acesso em: 18 de jul de 2013.

CAMARGO, Ana Maria de Almeida; GOULART, Silvana. **Tempo e circunstância: a abordagem contextual dos arquivos pessoais-procedimentos metodológicos adotados na organização dos documentos de Fernando Henrique Cardoso**. São Paulo: Instituto Fernando Henrique Cardoso, 2007.

CARMONA MENDO, Concépcion. Consideraciones sobre el concepto de documento privado**. "SIGNO. Revista de Historia de la Cultura Escrita"**, v. 3, p.11-23, 1996, Universidad de Alcalá de Henares. Disponível em: <http://dspace.uah.es/dspace/bitstream/handle/10017/7489/consideraciones_mendo_SIGNO_1996.pdf?sequence=1> Acesso em: 12 de jun de 2014.

CARMONA MENDO, Concepción. Consideraciones sobre el método en archivistica. In: **Documenta & Instrumenta**. Madrid, Universidad Complutense de Madrid, v.1, p. 35-46, 2004. Disponível em: <http://revistas.ucm.es/index.php/DOCU/article/view/DOCU0404110035A/19190> Acesso em: 11 de jun de 2014.

CONDE VILLAVERDE, Maria Luísa. La investigación em los archivos. Evolución de su contexto y contenido. **ARBOR**, v.CLXXXII, n.717, p. 33-37, 2006. Disponível em: <http://arbor.revistas.csic.es/index.php/arbor/article/view/5/5> Acesso em: 01 de dez. de 2013.

COOK, Terry. Arquivos Pessoais e Arquivos Institucionais: para um Entendimento Arquivístico Comum da Formação da Memória em um Mundo Pós- Moderno. **Revista Estudos Históricos**, v. 11, n. 21, p. 129-150, 1998 Disponível em:<http://bibliotecadigital.fgv.br/ojs/index.php/reh/article/view/2062/1201>Acesso em: 11 de ago de 2013.

FONSECA, Gabrieli Aparecida da. **A contribuição da identificação documental em arquivos pessoais: um estudo de caso no Instituto Fernando Henrique Cardoso**. Marília, 2015.

RODRIGUES, Ana Célia. **Identificação: uma nova função arquivística.** Revista EDICIC, v,1, n.4, p. 109- 129, 2011. Disponível em: <http://www.edicic.org/revista/> Acesso em: 08 de abr de 2013.

Análise documental de conteúdo e o percurso gerativo de sentido: na representação do documento de arquivo

Gilberto Gomes Cândido
Universidade Estadual Paulista
ggcandido@marilia.unesp.br

João Batista Ernesto de Moraes
Universidade Estadual Paulista
jota@marilia.unesp.br

Deise Sabbag
Universidade de São Paulo
deisesabbag@usp.br

1 Introdução

Muitas discussões são levantadas sobre a representação do conteúdo informacional dos documentos de arquivo, pois, tanto a Norma geral internacional de descrição arquivística (ISAD (G)), quanto a Norma internacional de registro de autoridade arquivística para entidades coletivas, pessoas e famílias (ISAAR (CPF)), não se comprometem, com a entrada ou a saída da informação no ambiente de Arquivo. Desta forma, a responsabilidade, passa a ser de cada instituição

mantenedora pela elaboração de seus instrumentos de pesquisa, como pela identificação de métodos que possam auxiliar nessa empreitada.

Com o propósito de contribuir com essa indagação, apresentam-se duas metodologias como subsídio de auxilio na representação do documento de arquivo e na identificação do seu contexto de produção: a) o Percurso Gerativo de Sentido auxiliando na análise da estrutura textual do documento; b) a Análise Documental de Conteúdo (ADC) por meio da etapa analítica, apresentando os pontos mais relevantes para a leitura do documento.

Na área de Ciência da Informação, essas propostas metodológicas vêm sendo utilizadas para auxiliar na identificação temática de textos de ficção e artigos científicos, com êxito por diversos autores, tais como: Moraes (2011; 2012); Moraes, Guimarães e Guarido (2007); Moraes, Damazo e Lara (2008); Moraes e Guimarães (2008); Sabbag (2012; 2013); Guarido e Moraes (2009); Moraes e Alves (2009); Garcia-Marco, Moraes, Garcia-Marco e Guimarães (2010); Moraes e Candido (2013); Candido, Lima e Moraes (2013) e Candido (2014).

Sendo assim, a pesquisa objetiva a aplicação dos métodos para a identificação do conteúdo do documento de arquivo. O *corpus* de análise é o documento de arquivo pertencente à subdelegacia de política de Ocauçu/SP, distrito desmembrado de Marília/SP em 1959, sob a responsabilidade do Centro de Documentação Histórica e Universitária de Marília – CEDHUM.

Posto isto, busca nas metodologias contribuições para melhorar a representação do documento de arquivo e a sua organização, disseminação, acesso e uso da informação.

2 Documento de arquivo

No entender de Jenkinson (1922), o documento para ser considerado e classificado como pertencente à classe de arquivo deveria ter sido elaborado em uma administração (pública ou privada), bem como transitado dentro dela (1922, p.11, tradução nossa).

Os autores Muller, Feith e Fruin, (1973, p.13) ampliam a definição ao considerarem, como documento de arquivo quaisquer “escritos, desenhos e material impresso, recebidos ou produzidos oficialmente por um determinado órgão

administrativo ou por um de seus funcionários, na medida em que tais documentos se destinavam a permanecer na custódia desse órgão ou funcionário" como prova de atividades.

Schellenberg (2012, p. 41) demonstra que, os documentos de arquivos são "[...] prova de suas funções, sua política, decisões, métodos, operações ou outras atividades, ou em virtude do valor informativo dos dados neles contidos".

Partilhando do entendimento de Schellenberg (2012), Rousseau e Couture (1998, p.137) afirmam que o documento de arquivo "é um conjunto constituído por um suporte e pela informação que ele contém, utilizáveis para efeitos de consulta ou como prova [...]".

Dessa forma, o valor de prova conferido ao documento de arquivo está em registrar atividades rotineiras de uma administração (pública ou privada) ou de indivíduos sociais, da mesma forma a proporcionar a identificação de quais foram as trajetórias realizadas para a sua elaboração, demonstrando ainda segundo os autores, suportes variados para se fazer o registro da informação.

Contudo, no entender de Jenkinson (1922), há também, a inferência de duas qualidades que asseguram ao documento de arquivo o valor de prova: imparcialidade atinente à verdade da ação administrativa; e a autenticidade sobre a manutenção, integridade e conservação da estrutura do documento de arquivo.

Complementando Jenkinson (1922), Duranti (1994) expõe mais três qualidades sendo consideras peculiares ao documento de arquivo e que possibilitam, também, manter o valor de prova: a naturalidade, a inter-relação/organicidade e a unicidade. A naturalidade do documento de arquivo provém de sua elaboração ininterrupta como reflexo das funções realizadas pelas instituições, de modo a provar as suas atividades; a inter-relação/organicidade se dá na relação que cada documento tem para com as funções que o geraram; "a unicidade provém do fato de que cada registro documental assume um lugar único na estrutura documental do grupo ao qual pertence no seu universo documental" (DURANTI, 1994, p.03).

As atividades de gestão de um arquivo devem garantir que essas cinco qualidades intrínsecas sejam atribuídas ao documento garantindo os atributos de especificidade, para que esses possam auxiliar na contextualização documental, e na manutenção do valor probatório dele recorrente.

Desse modo, para análise do documento de arquivo buscou-se na Análise Documental, em sua etapa analítica, meios para a realização da leitura de modo a ressaltar os trechos de seu conteúdo.

3 Análise documental

A Análise Documental (AD), no âmbito da Ciência da Informação, segundo os autores Gardin et al (1981); Garcia Gutierres (1984); Cunha (1987); Pinto Molina (1989); Coyaud (1966); Chaumier (1971;1974); Clauso Garcia (1933); Ruiz Perez (1992); Guimarães (2003); Pinto, Gavez (1999) e Pinto Molina (1991), é um conjunto de procedimentos que visam à identificação do conteúdo do documento, e à posterior síntese, tendo como finalidade a representação por meio de um documento secundário (produtos: catálogo, vocabulário controlado, etc.) considerado instrumento provedor de acesso, consulta e disseminação da informação.

Vale ressaltar que a Análise Documental se divide em dois processos: Análise Documental de Forma (ADF), processo descritivo dos elementos/caracteres extrínsecos ao documento; e Análise Documental de Conteúdo (ADC), processo de identificação dos elementos/caracteres intrínseco ao documento (PINTO MOLINA, 1992; RUIZ PÉREZ, 1992; FOX, 2005; NASCIMENTO, L. M. B. 2009).

Para este estudo se utilizará apenas os procedimentos da Análise Documental de Conteúdo (ADC), sendo esta dividida em duas etapas: 1) analítica e 2) sintética. A primeira corresponde aos processos de identificação da tipologia documental, leitura técnica e identificação de conceitos; a segunda, a seleção de conceitos, condensação documentária e representação documentária.

Dentre as etapas apresentadas, foca-se principalmente na etapa analítica do processo da ADC, mais especificamente na leitura técnica, em que o analista explora a estrutura textual do documento, identificando as partes mais relevantes do conteúdo.

Este procedimento de leitura compreende um conjunto de estratégias cognitivas e metacognitivas. Essas estratégias estão relacionadas com os objetivos que serão refletidos no decorrer da leitura documentária, dos quais seus aspectos são abordados por Pinto Molina (1993, p. 162-163), quando enfatiza que “as estratégias são desenvolvidas durante o período de leitura e não dependem somente do leitor e do texto, mas também dos objetivos documentários”. Essas operações são ações

mentais que agem de formas organizadas, e podem ser evidenciadas em duas categorias. A primeira, “cognitiva, que compreende a interpretação como comportamento automático e inconsciente; a segunda, a metacognitiva, que supõem atividades desautomatizadas, em que o leitor tem a consciência de como está lendo” (PINTO MOLINA, 1993, p. 162-163).

Deste modo, os recursos de leituras contêm características conceituais linguísticas e cognitivas, sendo utilizados no processamento da informação por meio das abordagens do *bottom-up* e *topdown,* que são considerados como procedentes da psicologia cognitiva.

Pinto Molina (1993, p.161) demostra que o *bottom-up* “é a leitura do tipo linear, indo das partes ao todo textual, que se baseia exclusivamente na estimulação visual”. O *top-down* é o método de leitura dedutivo, “em que se caminha de forma inversa, do todo para as partes, aproveitando-se do ‘conhecimento prévio’ ou ‘esquemas’”.

Dessa forma, com base na autora, entende-se que os métodos de leitura são dinâmicos e interpretativos; sua compreensão se estabelece pelas ações de significação, de forma a adequar a análise semântica e a análise de conteúdo.

4 Percurso gerativo de sentido

O Percurso Gerativo de Sentido resulta dos estudos de Greimas (1917-1992), que buscou nas estruturas textuais meios para explicar o desenvolvimento e o sentido do texto escrito.

Esse percurso é compreendido como uma expressão teórica da semiótica, que busca em seus componentes meios de articulação que promovam sua interação, de modo a apresentar uma *perspectiva de geração*, ou seja, pressupondo que todo objeto semiótico pode ser definido de acordo com a sua maneira de produção, deste modo os elementos que permeiam esse processo se correlacionam uns com os outros formando um ‘*percurso’* “que vai do mais simples ao mais complexo, do mais abstrato ao mais concreto” (GREIMAS; COURTÉS, 2008, p. 232).

Segundo Fiorin (1999, p.17), o percurso gerativo de sentido possui três níveis de estruturas, sendo: as fundamentais, as narrativas e as discursivas. A estrutura fundamental ou profunda é entendida como categoria semântica; a narrativa, como estrutura canônica; e a discursiva, como concretização de temas e figuras. Os

métodos de análise dos três níveis são distintos entre si, porém, ocorre uma correlação entre esses, para que tenha o pleno entendimento do texto, partindo da análise simples para a mais complexa.

Exemplificam-se, a seguir, os níveis:

- *Fundamental ou Profundo*: se estabelece por uma relação de contrariedades, de acordo com Fiorin (1999, p. 10 “[...] compreende a(s) categoria(s) semântica(s) que ordena(m), de maneira mais geral, os diferentes conteúdos do texto. Uma categoria semântica é uma oposição tal que *a* vs b”. No documento de arquivo isso pode ser observado por termos que apresentam esse nível contrariedade, exemplo: Deferido x Indeferido, Ratificar x Retificar, etc., com esta análise permite-se identificar traços simples e abstratos do texto, como sua oposição semântica;
- *Narrativo*: ocorre por meio da interlocução entre o sujeito e objeto que se conectam para serem reconhecidos narrando um fato/história. Destaca-se que se entende o documento de arquivo como sendo descritivo, ou seja, objetivo; ao descrever uma função e atividade. Neste nível temos a sequência canônica do texto: **Manipulação** (querer ou dever), **Competência** (saber ou poder fazer), **Performance** (transformação causada por uma mudança de um estado a outro) e **Sanção** (distribuição de prêmios e castigos);
- *Discursivo*: aponta os elementos concretos do texto por meio dos termos que se referem a **temas** e **figuras**, o *tema* contém características predicativas e interpretativas do mundo natural, e a *figura* simula o mundo natural por meio descritivo.

Entretanto, para que ocorra a estruturação dos percursos temático e figurativo, é indispensável a utilização da *isotopia* que, de acordo com Greimas e Courtés (2008, p. 278), “[...] constitui um crivo de leitura que torna homogênea a superfície do texto, uma vez que ela permite elidir ambiguidades”. Segundo os autores, é por meio das isotopias que se identificam as sequências semânticas que converte o texto lido num conjunto.

Complementando essa prospecção, Moraes e Guimarães (2006, p.07), demonstram que na etapa das "*estruturas discursivas*, os resultados obtidos nesta

fase devem remeter à etapa inicial, ou seja, às *estruturas fundamentais"*, isto é, o percurso lógico da construção do texto coerente.

Destaca-se que, no método Percurso Gerativo Sentido utilizou se apenas das estruturas narrativas e discursivas, em conjunto com AD: a primeira por meio da extração da sequência canônica do texto; a segunda possibilitando a análise dos temas e figuras de cada sequência.

5 Aplicação do percurso gerativo de sentido na estrutura textual do documento de arquivo

A aplicação do percurso gerativo de sentido foi sobre o documento de arquivo, escolhido de forma aleatória no Centro de Documentação Histórica e Universitária de Marília – CEDHUM, o documento pertencente à subdelegacia de polícia do distrito de Ocauçu/SP, distrito desmembrado de Marília/SP em 1959.

Realizou-se no documento a leitura documental, que possibilitou a identificação do nível narrativo e a extração dos elementos textuais presentes na sequência canônica: manipulação, competência, performance e sansão.

Posterior à identificação da sequência canônica, ocorreu a verificação do nível discursivo, compilando os temas e as figuras que apontavam os elementos concretos do documento.

A seguir apresentamos o experimento:

MANIPULAÇÃO
Em obediência ao oficio 1368 do sr. Dr. Delegado adjunto, dessa Delegacia Regional de Polícia, formulei um pequeno inquérito sobre a questão da novilha que fora indebitamente apropriada por José Oliveira, a qual é propriedade do sr. Pedro dos Santos.

Quadro 1- Manipulação – Elemento da Estrutura Canônica.
Elaborado: pelos autores

Quadro 2- Percurso Figurativo e Percurso Temático

TEMAS	**FIGURAS**
	Dr. Delegado
Inquérito	Delegacia Regional de Polícia
	Novilha
Apropriada	José Oliveira
Propriedade	Pedro dos Santos

Elaborado: pelos autores

Quadro 3 - Competência – Elemento da Estrutura Canônica.

COMPETÊNCIA
No correr dos interrogatórios, nada pude apurar e nem conseguir um entendimento entre ambos. Conforme se verifica do depoimento de José Santini fls.5 e a declaração de seu genro João de Oliveira, fls. 4, há certa contradição, pois João de Oliveira, proprietário da fazenda, confessa ter ali ficado uma novilha de propriedade de Pedro Bernardo, quando este de lá retirou o gado, ao passo que, Santini, campeiro da fazendo de seu genro, nega ter ali ficado qualquer gado de propriedade de Bernardo. O depoimento das testemunhas são todas a favor de Pedro Bernardo com exceção de José Ribeiro de Paulo, que vendeu-lhe o gado em questão.

Elaborado: pelos autores.

Quadro 4- Percurso Figurativo e Percurso Temático

TEMAS	FIGURAS
Interrogatórios	
Depoimento	José Santini
Declaração	João de Oliveira
	Fazenda
Confessa	Novilha
Retirou	Gado
	Pedro Bernardo

Elaborado: pelos autores.

Quadro 5 - Performance – Elemento da Estrutura Canônica.

PERFORMANCE
Nessas condições, envio os depoimentos das testemunhas para essa Regional, afim-de-que a caso seja aí apurado convenientemente.

Elaborado: pelos autores.

Quadro 6 - Percurso Figurativo e Percurso Temático

TEMAS	FIGURAS
Depoimentos	Testemunha
Apurado	Regional

Elaborado: pelos autores.

Quadro 7 - Sanção– Elemento da Estrutura Canônica.

SANÇÃO
Conforme se verifica das declarações de Pedro Bernardo e de diversas testemunhas, há causa de 2 meses uma vaca de Pedro Bernardo, passou para as invernadas de João de Oliveira, e lá foi ferrada com a marca da fazenda, e foi restituída a Pedro Bernardo.

Elaborado: pelos autores.

Quadro 8 - Percurso Figurativo e Percurso Temático

TEMAS	FIGURAS
Declarações	Pedro Bernardo
Diversas	Testemunhas,

Causa	Vaca
Ferrada	João de Oliveira
Marca	Fazenda

Elaborado: pelos autores.

Nos procedimentos de classificar temas e figuras, obteve-se como resultado a estrutura do percurso temático e do percurso figurativo, para o estabelecimento de temas.

Observa-se que as figuras e os temas, ao serem analisados de forma individual, não apresentam relevância, mas no momento em que a análise conceitual ocorre, constata-se o funcionamento e a construção do texto com o apoio da questão da isotopia, como sendo, um procedimento de construção do sentido sobre a dimensão sintagmática, que controla os significados das palavras diante do contexto.

Sendo assim, a análise do texto apresentou como traço sêmico o termo “Inquérito”, que tende a se reproduzir por meio da isotopia. Logo, o Percurso Gerativo de Sentido nos auxilia no entendimento do conteúdo textual do Documento de Arquivo, *posto que tal técnica fraciona o texto em trechos que, com a interpretação, proporciona a identificação do circuito narrativo, que o relaciona com a função que o elaborou para uma determinada atividade.*

Entretanto, ressalta que, tal método aplicado trabalha com o conteúdo textual ou resíduo desse independe de seu suporte. Desta forma, acredita que esse método não seria usual, em espécies documentais, tais como: nota de empenho, relatório de remessa, balancetes etc.

6 Considerações

O Documento de Arquivo é descritivo e objetivo, contudo, em decorrência de sua representação, esse também pode ser percebido como tendo um teor narrativo, no qual se percebe um grau de subjetividade existente pela intenção de interpretação, devido a isso o arquivista passa a ser um contador de história.

Os autores Duff e Harris (2002, p. 276) destacam que, no momento da representação do documento de arquivo, são selecionados determinados elementos/caracteres, que em seguida poderão ser utilizados para representá-lo. No entanto, “neste processo, não é a análise, listagem, reprodução, e assim por diante,

mas o seu principal meio é narrativa", ou seja, o ato de contar história (DUFF; HARRIS, 2002, p. 276, tradução nossa).

Portanto, o Percurso Gerativo de Sentido se apresenta de forma pontual para representação da informação do documento de arquivo, auxiliando na identificação da função e atividade, de modo a proporcionar uma melhor representação, disseminação, acesso e uso da informação.

Porém, também se expôs que o Percurso Gerativo de Sentido, em determinadas espécies documentais, tais como: nota de empenho, relatório de remessa, balancetes etc., não seria usual.

Referências

DUFF, W. M.; HARRIS, V. Stories and Names: Archival Description as Narrating Records and Constructing Meanings. **Archival Science**, v. 2, n. 3-4, p. 263-285, 2002.Disponivel em: < http://yalearchivalreadinggroup.pbworks.com/f/Duff&Harris.pdf>. Acesso em: 15 de Jul. 2014.

DURANTI, L. The concept of appraisal and archival theory. The American archivist. Chicago: **The Society of American Archivists**, v.57, n°2, 1994a, p.328- 344. Disponível em: < https://courses.ischool.utexas.edu/Winget_Megan/2007/Summer/Becker/Duranti_Concept_of_Appraisal.pdf> Acesso em: 16 Nov. 2013.

FIORIN, J. L. **Sendas e veredas da semiótica narrativa e discursiva.** *DELTA*. fev./jul. 1999, vol.15, no.1. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44501999000100009&lng=pt&nrm=iso. >. Acesso em: 16 Jan. 2014.

FLOCH, J. M. **Documento de Estudo do Centro de Pesquisa Socios semióticos.** Periódicos I. São Paulo: Centro de Pesquisas Socios semióticos, 2001.

FONSECA, M. O. **Arquivologia e Ciência da informação.** Rio de Janeiro: FGV, 2005. p. 121.

FOX, V. **Análisis documental de contenido:** principios y prácticas. Buenos Aires: Alfagrama, 2005.

FUJITA, M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. **DatagramaZero:** revista de Ciência da Informação, Rio de Janeiro, v. 5, n. 4, ago. 2004. Disponível em: <http://www.dgz.org.br/ago04/Art_01.htm>. Acesso em: 12 Mai. 2014.

GREIMAS, A. J; COURTÉS, J. **Dicionário de Semiótica**. São Paulo: Cultrix,2008.

GUIMARÃES, J. A. C. **Análise documentária em jurisprudência: elementos para uma metodologia de indexação de acórdãos trabalhistas brasileiros.** 1994. 250 f. Tese (Doutorado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1994.

JENKINSON, H. **A manual of archive administration:** including the problems of war archives and archive making. London, Edinburgh, New York, Toronto, Melbourne and Bombay: Oxford at the Clarendon Press, 1922. 281 p. Disponível em: <https://ia700302.us.archive.org/0/items/manualofarchivea00jenkuoft/manualofarchivea00jenkuoft.pdf>. Acesso em: 18 Dez. 2013

MORAES, J. B. E.; GUIMARÃES, J.A.C. Análisis documental de contenido de textos literarios narrativos: en busca del diálogo entre las concepciones de aboutness/meaning y de recorrido temático/recorrido figurativo. **Scire** (Zaragoza), 2006.

NASCIMENTO, L. M. B. **Análise documental e análise diplomática:** perspectivas de interlocução de procedimentos. 2009. 198 f. 2009. Tese de Doutorado. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista, São Paulo.

PINTO MOLINA, M. **El resumen documental:** principios y métodos. Madrid; Salamanca: Fundación Germán Sánchez Ruipérez; Madrid: Pirámide, 1992.

PINTO MOLINA. M. **Análisis documental: fundamentos y procedimientos**. 2. ed. rev. aum. Madrid; Eudema, 1993.

ROUSSEAU, Jean-Yves; COUTURE, C. **Os fundamentos da disciplina arquivística**. Lisboa: Publicações Dom Quixote, 1998.

RUIZ PÉREZ, R. **El análisis documental:** bases terminológicas, conceptualización y estructura operativa. Granada: Ed. Universidad de Granada, 1992.

SCHELLENBERG, T. R. **Arquivos modernos: princípios e técnicas**. Tradução de Nilza Teixeira Soares. Rio de Janeiro: Editora FGV, 2012, p.320.

Contribuições do percurso gerativo de sentido para a leitura documentária de textos narrativos de ficção

Deise Sabbag
Universidade de São Paulo
deisesabbag@usp.br

João Batista Ernesto de Moraes
Universidade Estadual Paulista
jota@marilia.unesp.br

1 Introdução

A leitura documentária é uma atividade importante dentro da etapa analítica da Análise Documental, possibilitando a identificação das partes mais relevantes do texto para que o analista possa realizar a identificação de conceitos. Nesse sentido, observamos que várias tipologias documentais oferecem uma estrutura textual que favorece a aplicabilidade de estratégias de leitura, como por exemplo, os documentos técnico-científicos que possuem os elementos pré-textuais, textuais e pós-textuais.

Observamos, também, que o emprego das mesmas estratégias de leitura para a análise de textos narrativos de ficção carecem de estudos para que essa tipologia narrativa não seja apenas representada por seu gênero ou por sua nacionalidade.

Neste sentido, este estudo teórico, de método bibliográfico, utilizou a teoria do Percurso Gerativo de Sentido para apresentar as regularidades subjacentes das narrativas de ficção que possibilitarão a representação do conteúdo dos documentos, tendo como objetivo a proposta de uma estratégia para a leitura documentária de textos narrativos de ficção.

O trabalho está organizado em três partes que apresentam ao leitor a definição de texto narrativo de ficção, seguido dos subsídios teóricos da área acerca a leitura documentária e, por fim, a proposta de leitura documentária de textos narrativos de ficção tendo como aporto teórico o Percurso Gerativo de Sentido.

O texto narrativo de ficção

Todo texto possui em sua construção uma estrutura superficial (microestrutura), uma estrutura profunda (macroestrutura) e um uso social (superestrutura). Para que um texto seja classificado como “texto narrativo”, torna-se necessário que sua microestrutura seja formada basicamente por elementos que indicam ações e mudanças de estado (verbos e indicadores de tempo e lugar); a macroestrutura apresenta personagens que sofrem mudanças de estado; e a superestrutura seja caracterizada por uma sequência de ações.

Neste sentido, apresentamos uma reflexão acerca do estabelecimento dos critérios que definem “o que é o texto narrativo de ficção desenvolvida a partir da literatura da área de Ciência da Informação, utilizando definições fornecidas por Lancaster (2004) e Beghtol (1994), fazendo uma justaposição como realizado por Moraes (2011).

Lancaster (2004) faz referência aos textos de ficção quando escreve sobre a indexação. O autor compara a indexação dos textos ficcionais em contrapartida dos textos não-ficcionais, e conclui que, essas duas tipologias textuais, apresentam questões diferentes para a indexação. Essas diferenças partem deste suas gêneses, pois as duas tipologias são construídas com objetivos diferentes, a ficção para entretenimento e a não-ficção para veicular informações.

Para Beghtol (1994), o mundo dos documentos pode ser divido em obras que nascem da imaginação de seus criadores e obras que nascem a partir da faculdade racional. Para a autora a ficção pode ser classificada como obra que surge da imaginação de seus criadores, sendo escrita em prosa narrativa, “narrativa pode ser tomada de forma ampla para incluir discursos que avançam de algum sentido de um ponto a outro” (BEGHTOL, 1994, p.7, tradução nossa).

Na análise das definições realizada por Moraes (2011), Lancaster (2004) apresenta uma perspectiva redutora sobre a obra literária, o que pode ser interpretado como desconhecimento em relação às manifestações literárias quando afirma que os textos de ficção destinam-se a entreter.

Evidentemente, essas obras podem servir a diversão, provocar emoções, mas defende-se a dimensão da ficção enquanto manifestação artística (presente e reconhecida desde a antiguidade) que não pode ser esquecida, muito menos desconsiderada na área de Ciência da Informação.

O reducionismo da definição de Lancaster (2004) é contraposto por uma amplitude na definição de Beghtol (1994) quanto ao texto de ficção, já que este é oriundo da imaginação de seus criadores. O texto literário encontra-se no limiar da imaginação, do imaginário, e como estabelecer os limites do 'o que é ou não imaginário'. Para a autora ficção é *narrative prose*, esta definição contempla também uma grande quantidade de textos que não são narrativos (ou apenas narrativos), mas não contempla textos que podem ser classificados como ficcionais.

Para construir uma definição de texto de ficção mais adequada aos estudos voltados para a área, Moraes (2011) toma como ponto de partida as afirmações de Beghtol (1994) sobre uma autocrítica a sua definição de ficção:

> É útil aceitar uma definição operacional por duas razões. Primeiro, nenhuma definição de ficção é universalmente aceita (...). Segundo, o exame de documentos de ficção para o armazenamento e recuperação de informação precisa incluir o maior número de exemplos de ficção possíveis (BEGTHOL, 1994, p. 07, tradução nossa).

O autor afirma que não há um consenso sobre o que seja ficção, apontando que uma definição deveria ser "tão abrangente de forma a abarcar todas as possibilidades do texto de ficção" (MORAES, 2011, p. 20).

Dessa forma, (MORAES, 2011) apresenta a ficção como uma forma de arte, uma manifestação artística construída com palavras que ultrapassam os escritos originários da imaginação do escritor, que pode conter elementos concretos da realidade.

Nessa perspectiva, o texto narrativo de ficção pode ser definido como uma obra de arte onde o artista é o escritor que cria criteriosamente os termos que irão representar o contexto escolhido pelo conteúdo semântico.

Considerar o texto narrativo de ficção como manifestação artística, como obra de arte, é lançar um olhar totalmente novo a esta manifestação textual, bem como para a leitura documentária e todo o processo de Análise Documental. É considerar que sua interação, e interlocução, atuam de um modo peculiar no desenvolvimento da competência comunicativa do analista.

Essa competência comunicativa proporcionará um tipo específico de interação com o leitor-analista: interação com uma criação artística; interação com uma forma artística; interação com uma expressão artística; proporcionando o conhecimento de sua tipologia textual e o estabelecimento de estratégias de leitura.

A leitura documentária

A leitura documentária compõe o processo de análise dos documentos. Este processo possui características e condicionantes particulares dentro da Análise Documental (AD).

Na literatura encontramos outras denominações para essa atividade: leitura profissional (MOURA, 2004), leitura técnica do documento (GUIMARÃES, 1994; DIAS; NAVES, 2007), leitura documentária (CINTRA, 1987; FUJITA, 2004), leitura para fins documentários (LARA, 1993), leitura rápida ou em diagonal (CHAUMIER, 1988), leitura atenta (CAVALCANTI, 1982), leitura profissional (FUJITA, 2004), misto de ler e passar os olhos (LANCASTER, 2004), exame do documento (NBR 12676), tomar atalhos (FOSKETT, 1973), leitor-bibliotecário (MOURA, 2004) etc.

A leitura documentária pressupõe uma modalidade específica de leitura onde requisitos técnicos serão utilizados para analisar uma variedade de documentos.

Kato (1985) desenvolve a concepção da legibilidade para a leitura do texto, levando em consideração a visão interativa da leitura, ou seja, abordando o ato de ler como uma interação entre o leitor e o texto.

Resumidamente, para Kato (1985), a legibilidade, em uma visão interativa, tem como função uma boa formação textual que leva em consideração não apenas a questão estrutural, mas também a funcionalidade; o conhecimento prévio do leitor por meio das operações de reconhecimento e, também, de inferências do novo por meio do conhecido; e o equilíbrio entre as estratégias cognitivas e metacognitivas.

Tomando como base a concepção de legibilidade de Kato (1985), Cintra (1987) concorda que os fatores elencados contribuem para a qualidade do texto, e apresenta, com base nos mesmos, quais estratégias de leitura seriam aplicadas para a prática profissional do bibliotecário.

Cintra (1987), Lara (1993) e Tálamo (1982) apropriam-se do conceito de leitura global utilizado por Grice(1975) para destacar a importância da compreensão do processo de leitura global para posterior compreensão da leitura documental.

Nesse processo de leitura global Grice (1982) apresenta o princípio de cooperação, ou seja, interlocução que ocorre entre o leitor e o texto, onde o leitor negocia com o texto até alcançar uma interpretação particular.

De acordo com Cintra (1987), na leitura para fins documentários o princípio de cooperação (autor/leitor), proposto com Grice (1982) é rompido, pois neste processo o autor não previu o leitor profissional.

Assim, como destaca Cintra (1987), considerar como fatores de legibilidade o conhecimento prévio e os tipos de estratégias que o texto exige, num processo interativo, é reconhecer que no processo de leitura estão envolvidos mais do que um conjunto de sinais visuais. Em realidade, é admitir que existem “aspectos cognitivos relacionados com conhecimentos armazenados na memória do leitor e comportamentos específicos durante a leitura” (CINTRA, 1987, p. 29).

Um conhecimento prévio muito importante para a leitura técnica do documento é o conhecimento da tipologia e das superestruturas textuais existentes em toda construção textual. Conhecendo a tipologia textual do documento o leitor pode dispensar a leitura palavra por palavra, pois ele domina as superestruturas, conseguindo interpretar as ideias centrais do texto. O domínio das superestruturas textuais permite que o leitor crie estratégias de leitura, ou, ações mentais que o levam a atingir seus objetivos (CINTRA, 1987, p. 31).

De acordo com Cintra (1987), a seleção de estratégias, durante a leitura, dependem de vários fatores: da finalidade da leitura, da experiência do leitor, da maturidade frente á tarefa de ler, do tipo de texto, da atenção mais concentrada em partes do texto, do grau de novidade do texto e da motivação de ler. Quanto às estratégias elas podem ser cognitivas (comportamentos automáticos e inconscientes) e metacognitivas (comportamentos desautimatizados, conforme o leitor tem consciência de como está lendo).

Para Cintra (1987), o leitor experiente seria caracterizado pelas habilidades de monitoramento, identificação das partes importantes do texto, concentração, segmentação do texto, ações corretivas de falhas detectadas no processo.

Sendo assim, a leitura técnica de qualquer documento, nas suas mais diversas formas de manifestações, permite a construção de substitutos, ou seja, de representações documentais.

Leitura documentária de textos narrativos de ficção

Tradicionalmente, na área de Ciência da Informação, aplicam-se, com pequenas alterações, os mesmos métodos de leitura documentária para uma enorme variedade de tipos de documentos. Para a leitura documentaria são utilizadas algumas estratégias, tais como: ler o título das obras; subtítulos; introdução e conclusão; frases introdutórias de parágrafos e capítulos; legendas de ilustrações; gráficos; tabelas; diagramas e suas explicações; conclusão; palavras ou grupos de palavras.

Esses procedimentos demonstram eficácia em documentos que oferecem uma estrutura textual, e física, adequada para a aplicabilidade dessas estratégias, como por exemplo, os artigos científicos, os relatórios técnicos, os livros didáticos etc. Mas quando as mesmas estratégias são aplicadas para a leitura documentária de um texto narrativo de ficção não se obtêm os mesmos resultados, ou seja, não se define o conteúdo temático do documento.

Uma solução seria a leitura integral dos textos narrativos de ficção, mas devido a questões de tempo o profissional não dispõe dessa possibilidade nas bibliotecas.

Dessa forma, qual recurso estratégico o analista deveria utilizar para a leitura técnica de um texto narrativo de ficção?

Sabe-se que o texto narrativo de ficção enquanto texto é composto por uma estrutura superficial (microestrutura), que compreende elementos linguísticos que auxiliam na sua composição; e por uma estrutura profunda (macroestrutura), onde existem os elementos que compõe a denominada semântica do texto. Também, por uma forma definida pelo uso social (superestrutura).

O texto narrativo de ficção possui quatro características que devem estar presentes em um texto para que ele seja uma narração: situações concretas, figuratividade, relações de posterioridade, concomitância e anterioridade entre os episódios relatados, e utilização preferencial do subsistema temporal do passado. Ele possui um plano de conteúdo que é concebido na forma de percurso gerativo. Isso confere ao texto narrativo de ficção a concepção de conteúdo organizado em articulações narrativas, ou seja, ele possui um plano para sua construção.

Considerando esse elemento, somado a noção da “figura do espetáculo” que é uma estrutura mínima, condensada, que pode ser a base para diversas narrações, os textos narrativos possuem uma constante estrutura de base; estrutura definida pelo Percurso Gerativo de Sentido.

Por meio do Percurso Gerativo de Sentido encontram-se os elementos necessários para o estabelecimento de estratégias para a leitura técnica de textos narrativos de ficção.

O Percurso Gerativo de Sentido possui patamares, sendo eles: fundamental; narrativo; discursivo e da manifestação. O nível narrativo é composto pela chamada sequência canônica, e é nesta sequência que encontramos os subsídios teóricos necessários para o estabelecimento de uma estratégia para a leitura documentária de textos narrativos de ficção.

A sequência canônica possui quatro fases descritas a seguir (FIORIN, 2011, p. 29-31):

- **Manipulação:** onde um sujeito age sobre outro para levá-lo a querer e/ou dever fazer alguma coisa. O sujeito é um papel narrativo e não, necessariamente, uma pessoa. São vários os tipos de manipulação, aqui descrevemos os quatro mais comuns: tentação, intimidação, sedução e provocação;
- **Competência:** onde o sujeito que realiza a narrativa é dotado de um saber e ou poder fazer;
- ***Performance*:** fase em que se dá a transformação central da narrativa (mudança de um estado a outro). Por exemplo: num conto de fadas em que uma princesa foi raptada pelo dragão, a *performance* será a libertação da princesa. Existe, portanto, um sujeito que opera a transformação e o sujeito que entra em conjunção, ou em disjunção com um objeto que pode ser distinto ou idêntico;
- **Sanção:** última fase onde há a constatação de que a *performance* se concretizou e o reconhecimento do sujeito que operou a transformação. A sanção pode ser cognitiva se há o reconhecimento que a competência se realizou; ou a sanção pode ser pragmática, com prêmios e castigos.

Assim sendo, pode-se considerar que a fase da **performance** e da **sanção**, geralmente, apresentam-se estruturalmente localizadas nos últimos capítulos das obras, seguindo a figura do espetáculo preconizada por Vladimir Propp. A figura do espetáculo poderia ser a base para diversas narrações, pois se apresenta como uma estrutura mínima que pode possibilitar diversas narratividades. Evidentemente que, ao longo de toda a narrativa ocorrem, também, situações de *performance* e sanção,

mas onde a **dimensão polêmica** da narrativa fica localizada é, a rigor, nos dois últimos capítulos. Portanto, geralmente, nos textos narrativos de ficção as situações de conjunção, ou disjunção, para os personagens principais, ocorrem nos capítulos finais. Evidenciando, também, a figura do espetáculo, ou seja, a presença de uma estrutura mínima que pode ser expressa com o enunciado “O pai dá um presente ao filho”.

Dessa forma, propõem-se, como estratégia de análise de textos narrativos de ficção, a leitura documentária dos últimos capítulos das obras, por ser uma parte significativa tematicamente na estrutura do texto. A aplicação desta estratégia possibilitará ao analista o emprego das categorias conceituais que permitirão a construção de enunciados de assunto.

Assim, para o exame de textos narrativos de ficção propõe-se que sejam utilizadas, as seguintes estratégias de leitura (leitura das seguintes partes do documento)

a) Título das obras;
b) Subtítulos;
c) Resumo;
d) Sumário;
e) Ilustrações, diagramas, tabelas e títulos explicativos (não são comuns em textos narrativos de ficção); e
f) Conclusão (**composta pelos últimos capítulos das obras**).

Considerações finais

A leitura documentária constitui uma atividade importante no bojo da fase analítica da Análise Documental. No processo de indexação, ela manifesta-se como uma etapa de fundamental importância para a compreensão dos textos que serão traduzidos e representados.

Essa modalidade de leitura exige determinados requisitos técnicos. Dentre os requisitos, destacamos o conhecimento prévio e os tipos de estratégias que o texto exige. Os conhecimentos prévios que apresentamos, neste trabalho, como importantes para a leitura de textos narrativos de ficção, foram os conhecimentos das tipologias e das superestruturas textuais.

Como tipologia, o analista (indexador) necessita reconhecer o texto narrativo de ficção como uma manifestação artística, caracterizada segundo o meio verbal, tendo seu objetivo artístico composto por palavras desvendadas pela capacidade de descodificação do leitor. Essa manifestação artística proporciona uma nova interação que desenvolve uma competência comunicativa no analista. Importante, também, que o analista reconheça, as superestruturas textuais dos textos narrativos de ficção, pois do reconhecimento das superestruturas que as estratégias de leitura serão desenvolvidas.

Essas estratégias foram alcançadas, neste trabalho, com base nos estudos da Teoria do Percurso Gerativo de Sentido que demonstrou que o texto narrativo de ficção possui um plano de conteúdo organizado, com regularidades narrativas subjacentes, ou seja, uma estrutura textual que permite ao analista identificar qual são as partes mais importantes do texto para a leitura documentária e, posterior, identificação de conceitos.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 12676**: métodos para análise de documentos: determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro, 1992.

BEGHTOL, C. **The classification of fiction**: the development of a system based on theoretical principles. Metuchen: Scarecrow Press, 1994.

CAVALCANTI, C. R. Indexação. **Estudos Avançados em Biblioteconomia e Ciência da Informação**, Brasília, DF, v. 1, n. 1, p. 211-233, 1982.

CHAUMIER, J. Indexação: conceito, etapas e instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun. 1988.

CINTRA, A. M. M. Estratégias de leitura em documentação. In: SMIT, J. W. (Coord.). **Análise documentária:** a análise da síntese. Brasília, DF: MCT; CNPq; IBICT, 1987. p. 28-35.

DIAS, E. W.; NAVES, M. M. L. **Análise de assunto:** teoria e prática. Brasília: Tesauros, 2007.

FIORIN, J. L. **Elementos de análise do discurso**. 15. ed. São Paulo: Contexto, 2011.

FOSKETT, A. C. A **abordagem temática da informação.** São Paulo: Polígono, 1973.

FUJITA, M. S. L. A leitura documentária na perspectiva de suas variáveis: leitor-texto-contexto. **DatagramaZero**: revista de Ciência da Informação, Rio de Janeiro, v. 5, n. 4, ago. 2004. Disponível em: <http://www.dgz.org.br/ago04/Art_01.htm>. Acesso em: 12 jan. 2012.

GRICE, H. P. Logic and conversation. In. COLE, P., MORGAN, P. **Syntax and semantics 3: speech acts**. New York: Academic Press, p. 41 – 58. Disponível em: < http://www.ucl.ac.uk/ls/studypacks/Grice-Logic.pdf >. Acesso em: 2 abril 2013.

GRICE, H. P. Lógica e conversação. In: DASCAL, M. **Fundamentos metodológicos da linguística**. Campinas: 1982. p. 81-103.

GUIMARÃES, J. A. C. Análise documentária no âmbito do tratamento da informação: elementos históricos e conceituais. In: RODRIGUES, G. M.; LOPES, I. L.(Org.). **Organização e representação do conhecimento na perspectiva da Ciência da Informação**. Brasília: Thesaurus, 2003. p. 100-117.

KATO, M. A. Uma visão interativa da legibilidade. **Ilha do Desterro**, Floiranópolis, n. 13, p. 57-66, 1985.

LANCASTER, F. W. **Indexação e resumos.** Brasília, DF: Briquet de Lemos, 2004.

LARA, M. L. G. de. **A representação documentária**: em jogo a significação. 1993. 132 f. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1993.

MORAES, J. B. E. de. **A questão do aboutness no texto narrativo de ficção:** perspectivas metodológicas para a Ciência da Informação. 2011. 81f. Tese (Livre-Docência) – Faculdade de Filosofia e Ciências de Marília, Universidade Estadual, Marília, 2011.

MOURA, M. A. Leitor-bibliotecário: interpretação, memória e as contradições da subjetividade. **Perspectivas em Ciência da Informação**, Belo Horizente, v. 9, n. 2, p. 158-169, jul./dez. 2004.

TÁLAMO, M. G. M. **Comunicação e funcionamento da linguagem**. 1982. Dissertação (Mestrado em Ciências da Comunicação) – Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1982.

TATIT, L. Abordagem do texto. In. FIORIN, J. L. (Org). **Introdução à linguística:** I: objetos teóricos. São Paulo: Contexto, 2007. p. 187-209.

O modelo teórico do percurso gerativo de sentido para a elaboração de resumos de textos científicos: uma releitura do produto informacional resumo

Roberta Caroline Vesú Alves
Universidade Estadual Paulista
robertavesu@gmail.com

João Batista Ernesto de Moraes
Universidade Estadual Paulista
jota@marilia.unesp.br

Introdução

A elaboração de resumos constitui em procedimento da Análise Documental de Conteúdo, tradicionalmente voltado para textos científicos, visando à representação do conteúdo do documento em texto sucinto (GUIMARÃES, MORAES, GUARIDO, 2007; MOREIRO GONZÁLEZ, 2004; PINTO MOLINA, 1993). Por isso, respeita as características desse tipo de texto, explicadas por teorias estruturais da Linguística Textual (MOREIRO GONZÁLEZ, 2004; VAN DIJK, 2000).

Contudo, verificou-se que o Percurso Gerativo de Sentido, proveniente da Análise do Discurso e Semiótica Greimasiana, também coopera para elucidar compreensão de algumas características do texto científico (FIORIN, 2008; GREIMAS; LANDOWSKI, 1979). Por isso, objetivou-se analisar a contribuição do modelo teórico Percurso Gerativo de Sentido para a compreensão da elaboração do produto informacional resumo.

Utilizou-se, para isso, a metodologia de estudo exploratório com revisão de literatura, para análise qualitativa das informações.

Verificou-se que o Percurso Gerativo de Sentido elucida o modo de expressão do saber científico em textos e a organização textual, que são elementos importantes a serem representados em resumos.

Resumos em análise documental de conteúdo

A Análise Documental de Conteúdo apresenta estudos sobre procedimentos para tratamento do conteúdo de diferentes documentos, mas, os estudos sobre elaboração de resumos estão amplamente voltados para os textos científicos.

Os procedimentos de Análise Documental de Conteúdo, segundo Guimarães, Moraes, Guarido (2007), consistem em análise, síntese e representação para fins de recuperação, mediante a etapa analítica (leitura de partes importantes com estratégias metacognitivas para identificação de conceitos) e etapa sintética (seleção, condensação em resumo e representação em índices).

A representação de conteúdo, segundo Lara (1993), ocorre por meio de condensação intensiva, gerando produtos documentais como os resumos, que apresentam semelhança com o texto de origem e seu sistema semiótico, portanto, semelhante com a estrutura e a temática do texto original.

O resumo consiste no produto da transformação do texto original em outro texto reduzido sobre o conteúdo de um documento, com a função de intermediar a informação para o usuário (PINTO, 2001; PINTO MOLINA, 1993).

O método de resumir, segundo Pinto (2001), consiste no sofisticado processo de redução sobre o conteúdo do texto e de construção textual, realizados cognitivamente para representar o texto original, também, apresenta um plano de organização superficial que representa o texto original. A superestrutura do texto original e a temática (macroestrutura) devem ser transferidas para o resumo (MOREIRO GONZÁLEZ, 2004).

Os procedimentos para elaboração de resumos, segundo Pinto Molina (1993), são complexos, mas, de modo geral seguem métodos de leitura, análise, síntese e representação. Nesse sentido, Pinto (2001, p.65, tradução nossa) diz que o processo geral de resumir compreende,

> [...] atividades cognitivas praticadas sobre objetos linguísticos mediante o emprego de ferramentas lógicas com um objetivo documental. O resultado é um resumo/produto, um breve texto representativo, intencional, não unívoco e com vocação de sinônimo. A singularidade intrínseca deste complexo operativo analítico/sintético, a diversidade tipológica documental, a personalização das demandas documentais e a variabilidade das condições de produção são, entre outros aspectos, fatores determinantes do grau de complexidade de tais operações.

Segundo Pinto Molina (1993, 2001), a leitura para análise do texto original identifica suas características fundamentais (forma, classe e estrutura da informação), bem como temas principais e secundários da estrutura cognitiva do documento e as seções importantes, com a essência do documento. Portanto, conforme a autora, ANSI/NISO Z39.14-1997 (2009) e Simões (2014), a leitura observa as partes dos textos científicos que compreendem: introdução, objetivos, metodologia, resultados e conclusão (IOMRC).

Nesse sentido, a leitura produz vários processos cognitivos de redução e condensação por meio de macrorregras, que leva à macroestrutura global (PINTO, 2001; VAN DIJK, 2000). As macrorregras mostram como ocorre o ato de resumir cognitivamente, considerando as informações pertinentes para uma tarefa, esquemas e objetivos do leitor para formar as macroestruturas (GUIMARÃES, 2013).

A síntese para fins de representação, segundo Pinto Molina (1993), compõe a informação resultante da análise em resumo, considerando a coesão e coerência, de modo fiel ao original, preciso, claro e conciso.

A elaboração de resumos também deve considerar o *aboutness*, ou seja, o conteúdo intrínseco e relativamente permanente, segundo Beghtol (1983). O *aboutness* pode ser identificado por meio do Percurso Gerativo de Sentido (MORAES, 2011). Por isso, apresenta-se a seguir os aspectos estruturais e temáticos desse modelo teórico e Semiótica Greimasiana, importantes para representação em resumo.

Percurso gerativo de sentido e semiótica do discurso científico para elaboração de resumos

O Percurso Gerativo de Sentido, segundo Moraes (2011), contribui para identificação do *aboutness* revelado pelas regras de composição do texto narrativo e ficcional. Isso também ocorre com outros tipos de textos devido à narratividade. Esta, segundo Fiorin (2008), consiste em uma narrativa mínima, que ocorre com a transformação de estado de algo (estado inicial, transformação e estado final), o que permite a análise da estrutura e sentido do texto científico, considerando também a semiótica do discurso científico estabelecida por Greimas.

O modelo teórico do Percurso Gerativo demonstra os patamares da produção e interpretação do sentido do texto. Os seus elementos estruturais ou sintáxicos e semânticos estão distribuídos em níveis que compreendem o profundo ou fundamental (termos opostos com algum traço comum), o narrativo (fases de manipulação, competência, performance e sanção), e o discursivo (temas, figuras e percursos), mediante isotopias (FIORIN, 2008).

O *aboutness* pode ser identificado em Análise Documental de Conteúdo, a partir desse modelo teórico, considerando as fases de manipulação e sanção na narrativa, formando o tema principal em estrutura profunda (MORAES, 2011). Também, essas fases da narrativa evidenciam o estado inicial e final do saber científico, importantes para a comprovação do *aboutness* do discurso científico.

Segundo Greimas e Landowski (1979), o discurso científico pode ser analisado por meio das dimensões cognitiva e pragmática. A primeira identifica a ideia de transformação de estado do saber científico (ideia inicial de *não-saber*, a transformação e o estado final de *saber*, com a conjunção com objeto de valor), sendo que a fase de manipulação expressa o *não-saber* algo e a sanção o *saber* alcançado, para compor o tema principal. Nesse contexto, a competência cognitiva corresponde ao conhecimento necessário para a busca do *saber* e a performance, o desenvolvimento da mudança de estado (GREIMAS; LANDOWSKI,1979).

Para isso, segundo Greimas e Courtés (2008), o sentido enunciado pelo pesquisador sobre o saber científico é expresso em dimensão pragmática (interior à dimensão cognitiva), conforme os programas narrativos (acontecimentos percebidos pelo enunciatário). Segundo Barros (2008), os programas narrativos apresentam o modelo: $PN=F[S_1 \rightarrow (S_2 \wedge O_v)]$ (F função/fazer; –> transformação; S_1 sujeito do fazer; S_2 sujeito do estado; ∧ conjunção; v disjunção; O_v objeto de valor).

Desse modo, pode-se identificar a busca do *saber* científico em dimensão cognitiva, por meio da dimensão pragmática, e as partes principais desse tipo de discurso para compor o resumo. Nesse sentido, entende-se que os elementos da pesquisa científica observados em Cervo e Bervian (2005), se relacionam às teorias Greimasianas da seguinte forma:

- Introdução: contexto que motivou a elaboração da pesquisa, incluindo de modo geral, o tema de pesquisa, problema, hipóteses, tese, motivações (ideia inicial de *não-saber*, que leva à busca do *saber* – manipulação);
- Objetivo: metas a serem alcançadas para comprovar hipóteses e proposições (ideia que norteia a transformação de estado e a busca do *saber* – manipulação visando à sanção);
- Metodologia: meios de coleta e análise das informações, técnicas para estabelecer a verdade (meio pelo qual o *saber* será alcançado; conhecimento que colabora com o modo de busca do *saber* – competência);
- Resultados: expõe como ocorreu a coleta e análise das informações experimentais e teóricas, bem como suas relações (desenvolvimento da mudança de estado ou transformação – performance);
- Conclusão: apresenta a análise das informações em relação aos objetivos propostos, comprovando ou contestando hipóteses, e solucionando o problema de pesquisa (estado final de *saber* – sanção).

Para demonstrar a construção de resumos por meio das teorias estudas foi escolhida a representação em resumo informativo estruturado, utilizando o texto de Queiroz (2015). Inicialmente, foi estabelecida a macroestrutura e *aboutness*, observando as fases de manipulação e sanção:

- Manipulação: a ideia inicial de *não-saber*, que leva a busca do *saber* foi identificada no problema de pesquisa e objetivo geral, pois, o estudo pretende analisar se os recursos de Tecnologia Assistiva disponíveis nas salas de recursos multifuncionais atendem a demanda funcional dos alunos com deficiência física.
    - PN=F[dúvida –>(recursos de Tecnologia Assistiva ∧ adequados aos alunos deficientes físicos)]
- Sanção: a Tecnologia Assistiva das salas de recursos multifuncionais como materiais pedagógicos, equipamentos e mobiliários atende parcialmente os alunos com deficiência física, pois, alguns alunos encontram dificuldades em seu uso e manipulação;
    - PN=F[saber –>(recursos de Tecnologia Assistiva ∧ parcial inadequação aos alunos deficientes físicos)].

Verificou-se no discurso pragmático que ocorreu uma conjunção parcial com o objeto na sanção. O discurso cognitivo reconhece a transformação do estado inicial de *não-saber* (dúvida) se a Tecnologia Assistiva atende os alunos com deficiência física para o estado final de *saber* que atende parcialmente, o que resultou no *aboutness* “adequação *versus* inadequação” ou, mais especificamente, “dúvida de adequação dos recursos de Tecnologia Assistiva para alunos deficientes físicos *versus* parcial inadequação dos recursos de Tecnologia Assistiva para alunos deficientes físicos”.

A partir disso, o resumo estruturado do texto de Queiroz (2015) foi elaborado considerando as fases de manipulação (introdução e objetivo), competência (metodologia), performance (resultados) e sanção (conclusão):

Introdução: a inclusão escolar permite teoricamente o acolhimento e o desenvolvimento de competências e habilidades de educandos com deficiências, mas, apresenta dificuldades na prática, que levam a indagar se a os recursos de Tecnologia Assistiva atendem a demanda de alunos com deficiência física. **Objetivo**: analisa se os recursos de Tecnologia Assistiva das salas de recursos multifuncionais atendem a demanda de alunos com deficiência física, acompanhados pelo Atendimento Educacional Especializado. **Metodologia**: triangulação das informações qualitativas e quantitativas coletadas pelos questionários School Function Assessment, para identificar o perfil funcional dos alunos, e Tecnologia Assistiva para Educação II, para análise dos recursos utilizados, e entrevista semiestruturada e não padronizada, aplicados com sete professores do Atendimento Educacional Especializado na Rede Municipal de ensino, que trabalham com alunos com deficiência física. **Resultados**: o perfil funcional dos quinze alunos com deficiência física estudados consiste em idade entre seis e quatorze anos, cursando o ensino fundamental, dez portadores de Paralisia Cerebral e cinco com Mielomeningocele, que usavam para sua mobilidade, cadeira de rodas manual, bengala canadense e três tinham deambulação independente. Além disso, seis utilizam a escrita manual, seis o computador, dois as letras móveis e um não escrevia, sendo que sete realizavam atividades próximas ao nível de sua série. A avaliação dos professores sobre a adequação dos recursos de Tecnologia Assistiva das salas de recursos multifuncionais às necessidades dos alunos com deficiência física apontou algumas dificuldades de mobilidade e acessibilidade, mobiliários inadequados, entre outros aspectos. **Conclusão**: os recursos de Tecnologia Assitiva como materiais pedagógicos, equipamentos e mobiliários atendem parcialmente os alunos com deficiência física, pois, alguns alunos encontram dificuldades em seu uso e manipulação.

Diante do exposto, entende-se que o processo de resumir também pode ser elucidado pelos elementos teóricos do Percurso Gerativo de Sentido e Semiótica do texto científico.

Considerações finais

O Percurso Gerativo de Sentido e a Semiótica Greimasiana contribuíram para elucidar o modo de expressão do saber científico em textos, seu sentido (*aboutness*) e as estruturas subjacentes de organização textual. Os elementos identificados a partir da teoria proposta contribuíram para o esclarecimento da representação em resumos, oferecendo uma releitura desse produto informacional de modo mais consciente sobre sua estrutura e temática.

O saber científico observado por meio da narratividade (estado inicial, transformação e estado final), dimensões pragmática e cognitiva do texto e Percurso Gerativo de Sentido, revelou a macroestrutura textual e *aboutness*, além das partes importantes para representação em resumo, que correspondem à superestrutura (introdução, objetivo, metodologia, resultados, conclusão).

Portanto, verificou-se que o modelo teórico do Percurso Gerativo de Sentido e a Semiótica Greimasiana atuam estrategicamente para o entendimento do sentido do texto científico e seus aspectos estruturais, importantes para o entendimento da elaboração do produto documental resumo.

Referências

ANSI/NISO Z39.14-1997(R2009). **Guidelines for Abstracts**. NISO. Baltimore: Maryland,2009.

BARROS, D.L.P.de. **Teoria semiótica do texto**. São Paulo: Ática,2008.

BEGHTOL, C. Bibliographic classification theory and text linguistics. **Journal of Documentation**, London, v.42, n.2, p.84-113, june.1986.

CERVO, A.L.; BERVIAN, P.A. **Metodologia científica**. Pearson: São Paulo,2005.

FIORIN, J.L. **Elementos de análise do discurso**. São Paulo: Contexto,2008.

GREIMAS, A.J.;COURTÉS, J. **Dicionário de Semiótica**. São Paulo: Contexto,2008.

GREIMAS, A.J.; LANDOWSKI, E. **Análise do discurso em Ciências Sociais.** São Paulo: Global,1979.

GUIMARÃES, E. **Texto, discurso e ensino**. São Paulo: Contexto,2013.

GUIMARÃES, J.A.C. O resumo como instrumento para a divulgação e a pesquisa científica. **Revista Brasileira de Educação Especial,** Marília, v.11, n.1, p.3-16, jan./abr.2005.

GUIMARÃES, J.A.C.; MORAES, J.B.E.de; GUARIDO, M.D.M. Análisis documental de contenido de textos narrativos. **Ibersid,** Zaragoza, v.1, p.93-100.2007.

LARA, M.L.G.de. **A representação documentária**. Dissertação (Mestrado em Biblioteconomia e Documentação)–ECA, USP, São Paulo,1993.

MORAES, J.B.E.de. **A questão do aboutness no texto narrativo de ficção**. Tese (Livre-Docência em Linguística e Documentação)–FFC, UNESP, Marília,2011.

MOREIRO GONZÁLEZ, J.A. **El contenido de los documentos textuales**. Gijón: Trea,2004.

PINTO MOLINA, M. **Análisis documental**. Madrid: Eudema,1993.

PINTO, M. **El resumen documental**. Madrid: FGSR,2001.

QUEIROZ, F.M.M.G.de. **Tecnologia assistiva e perfil funcional dos alunos com deficiência física nas salas de recursos multifuncionais.** Dissertação (Mestrado em Educação)–FFC, UNESP,Marília,2015.

SIMÕES, M.daG.deM. **Resumo.** Coimbra: Universidade de Coimbra,2014. Slide da disciplina “Resumo documental” para PPGCI–Unesp.

VAN DIJK, T.A. **El discurso como estructura y proceso**. Barcelona: Gedisa,2000.

El control de vocabulario y su aporte a la investigación: el caso de las ciencias agrarias de Uruguay

Lucía Simón
Universidad de la República
lsimon@csic.edu.uy

Mario Barité
Universidad de la República
mario.barite@fic.edu.uy

1 Introducción

La política explícita en Ciencia y Tecnología (C&T) en Uruguay tiene -desde el año 2005- un desarrollo creciente reflejado en la inversión que el país realiza en este rubro, sobre todo, a partir de la creación de la Agencia Nacional de Investigación e Innovación (ANII), cuya misión es ejecutar los lineamientos políticos y estratégicos del Estado en Investigación e Innovación (Bértola, et al, 2005). En la orbita de esta Agencia, se destaca la creación -en el año 2007- del Sistema Nacional de Investigadores (SNI) con el objetivo principal de fortalecer y expandir la comunidad científica. La determinación de impulsar a las entidades que trabajan en C&T en el Uruguay, no es ajena a las necesidades del país por lo que el Plan Estratégico Nacional en Ciencia, Tecnología e Innovación (PENCTI) establece sectores y núcleos prioritarios -Salud Humana y Animal / Producción agropecuaria y agroindustrial integran la nómina compuesta por seis sectores adicionales- a fortalecer que deben ser tenidos en cuenta por la ANII y el SNI (PENCTI, 2010).

El presente trabajo se propone evidenciar cómo la investigación científica en Ciencias Agrarias en Uruguay, responde a estos lineamientos a través del estudio de la producción científica de los investigadores de más alto nivel en el área identificados en el SNI en la categoría III. Para acceder a esta categorización es necesario tener formación de doctorado, producción de conocimiento original, reconocimiento

internacional, creación y dirección de grupos de investigación y creación de capacidades para la investigación -institucionales y de formación de investigadores.

Si bien, el SNI brinda acceso a los currículum *vítae* -CV- de sus investigadores, estosaportan una visión parcial sobre las temáticas de investigación ya que la falta de normalización no permite evidenciar de forma clara los principales temas abordados.

En este sentido, se cree que la aplicación de técnicas propias de la Organización del Conocimiento como el control de vocabulario son apropiadas para mapear y el avance de las ciencias y disciplinas en general puesto que el control de vocabulario comprende un

> conjunto de técnicas y procedimientos que se aplican sobre el lenguaje para resolver problemas de comprensión, ambigüedad, alcance y relación entre términos que expresan conceptos, que se fundamenta en la idea de que el lenguaje natural, tal como se expresa en textos y documentos de la más variada especie genera -inclusive en comunidades de usuarios especializadas y más o menos sólidas y cerradas- dificultades de interpretación y de comunicación, las que son vistas como desviaciones del discurso metódico y preciso (Barité, 2014).

Adicionalmente, se agregan dos capacidades de esta herramienta de importancia sustantiva. La primera, relacionada con su

> capacidad para mejorar la eficacia y la eficiencia de sistemas de información (**National Information Standards Organization,** 2010, p. 1), pero también de sistemas de navegación web y de otros sistemas que se construyen a partir de la utilización del lenguaje para describir la realidad. En esta línea el principio de consistencia, entendido como la “coherencia en la aplicación de principios, reglas y procedimientos en lo relativo a la creación y uso de estructuras conceptuales o lenguaje natural para fines de clasificación o indización” (Barité et al., 2013), guarda una relación directamente proporcional con el control de vocabulario, pues se presume que a mayor control de vocabulario, mayor consistencia, y por ende, mayor eficiencia del sistema de información, en cuanto logra facilitar la descripción temática, la búsqueda y la recuperación (Barité, 2015).

La segunda -aplicada en este trabajo- es su

> aptitud para organizar campos conceptuales, áreas temáticas y disciplinas a través de la destilación de los conceptos mayormente asociados entre sí en la literatura, sea que estén claramente establecidos y aceptados, o sea que se encuentren en cuestión. De este modo, los vocabularios controlados ofrecen una placa fija del estado del arte de una especialidad en un momento determinado y desde una determinada perspectiva, considerando tanto las prácticas, las teorías y los tópicos ya consolidados como sus puntos de ruptura y conflicto. En su evolución esa placa se transforma en móvil,

> exhibiendo los cambios producidos dentro de un campo temático a lo largo del tiempo (Barité, 2015)

El presente estudio, constituye una primera aproximación exploratoria sobre la aportación de la herramienta a la fijación de los criterios aportados por la política en C&T en Uruguay para evaluar su pertinencia en función de los resultados que se aprecian en el SNI. Para el caso estudiado, esto puede constituir un aporte especialmente importante en -al menos- dos sentidos. Por un lado, el crecimiento económico de Uruguay depende en gran medida del desarrollo agropecuario (Bértola, *et al., 2005*) por lo que el correcto mapeo de la situación sobre investigación en esta área hace pensar que redundará en la mejor utilización de los recursos económicos. En segundo lugar, el avance en cualquier disciplina requiere de estados del arte relativamente sólidos que permitan guiar el camino a seguir.

2 La investigación en Ciencias Agrarias en el Uruguay: descripción basada en el control de vocabulario

La idea de que el control de vocabulario pueda considerarse como una herramienta válida para evidenciar el estado del arte de cualquier disciplina y su eventual concordancia con los lineamientos nacionales en materia científico tecnológica, se sustenta en la aplicación de las funciones básicas del control de vocabulario que se desprenden del análisis comparado del contenido de tres de los estándares de mayor prestigio en el mundo (*British Standards Institution*, 2005; ***National Information Standards Organization*, 2010;** *International Standard Organization*, 2011).

En el presente trabajo, se testea la hipótesis de que el control de vocabulario puede aportar una visión sobre los temas nucleares de la investigación en Ciencias Agrarias de Uruguay.

Se recolectaron los datos relativos a la producción científica en revistas científicas arbitradas de los investigadores presentes en el Nivel III del SNI correspondientes al área Agraria. Debido a que el formato del CV no prevé de forma obligatoria la consignación de palabras clave, se accedió a los artículos -o resúmenes- para extraer los términos consignados originalmente por los investigadores.

De los 451 artículos extraídos del CV, el 69,4% presentaron 1.210 palabras clave. El resto de los artículos no contenían esta descripción-. La tarea de control de vocabulario sobre las 1.210 palabras clave propias del lenguaje natural se realizó en

tres etapas donde se depuraron los términos para alcanzar un nivel de detalle que evidencie las áreas principales de investigación. En los tres casos, se tomaron como base las cinco funciones básicas del control de vocabulario -restricción, desambiguación, normalización, jerarquización y relación- con diferentes niveles de detalle (Barité, 2014).

La primera etapa consistió básicamente en el control de equivalencias, de sinónimos y variantes cumpliendo con la función de *normalización*. En la segunda, se aplicó también la función de *restricción,* mientras que, en el último caso, se aplicaron también las funciones de jerarquización y relación. La función correspondiente a *desambiguación* fue utilizada en menor medida, pues el área aporta *per se* los términos ya desambiguados por su alta especificidad.

En la tabla I se aportan los datos correspondientes a los términos controlados en cada etapa.

Tabla I: Etapas de control de vocabulario

	Inicio	Etapa I	Etapa II	Etapa III
Palabras clave	1210	665	147	19

Fuente: Elaboración propia

Etapa I: *Control de equivalencias, control de sinónimos y variantes, normalización.*

Consistió en la aplicación de la función de *normalización* del control de vocabulario en el entendido que la misma comprende la fijación de forma y de sentido, y la situación ponderada de cada término en la estructura conceptual de pertenencia, lo que tiende a una mayor coincidencia entre la indización y las búsquedas de los usuarios (Aitchison, Gilchrist & Bawden, 2000). Esta función se precisó en dos niveles. Por un lado, se realizó el control de equivalencias sobre las 1.210 palabras clave que fueron traducidas del idioma inglés. En segundo lugar, se controlaron los sinónimos y variantes y se mantuvieron solo aquellas palabras clave que se consideraron preferidas en función de la garantía literaria.

De la depuración realizada se excluyeron 545 términos -474 correspondientes al control de equivalencias y 71 al control de sinónimos y variantes- lo que redujo el campo de estudio a aun total de 665 términos.

Etapa II: Restricción, normalización

Consistió en la aplicación de la función de *restricción*. Barité (2015) menciona cuatro aspectos sobre los cuales trabaja la restricción: cantidad de términos, extensión de los términos, alcance conceptual y relaciones. En esta investigación se operó sobre la restricción de la extensión de los términos y se excluyeron aquellas palabras claves sintagmáticas con una estructura igual o superior a cuatro palabras. En caso de que superaran esa extensión y el nivel de incidencia fuera significativo se normalizó la expresión a una con menor cantidad de palabras.

Cumplida la depuración en base a la extensión, se realizó nuevamente la operación de normalización efectuada en la Etapa I. Como resultado, se excluyeron 415 por contener tres o más palabras y se normalizaron 103. El resultado final de esta etapa fue la exclusión de 518 palabras clave.

Etapa III: Jerarquización, relación y normalización.

En este punto el universo de palabras se redujo a 147. Se aplicaron las funciones de jerarquización, relación y normalización (Barité, 2015) lo que permitió excluir un total de 128 palabras clave.

Las relaciones abarcan un total de 83 términos que se excluyeron del universo de estudio. El criterio consistió en la identificación de términos autorizados y no autorizados y la exclusión de los últimos fue determinada por el principio de garantía literaria que permitió establecer la existencia del término, la correspondencia del mismo con un concepto reconocido por especialistas, la representatividad del mismo frente a eventuales sinónimos y la garantía literaria que lo respalda. La función de jerarquización se trabajó desde las relaciones intra-término, donde se dilucida la situación de sinónimos, cuasi sinónimos y variantes; y las relaciones entre términos, que tradicionalmente se establecen bajo distintas modalidades de relación jerárquica o asociativa (Cintra *et al.*, 2002, Aitchison & Clarke, 2004, Naumis, 2007). Sólo aquellos términos que representaban asociaciones con frecuencias relativamente altas fueron consignados. Este proceso excluyó a 18 términos. Como en las etapas precedentes, se normalizaron nuevamente los términos por lo que se excluyó un total de 128 términos.

El uso de las funciones del control de vocabulario permitió pasar de un universo de 1210 términos a uno de 14. En la tabla II se muestran los dos términos nucleares que se determinaron y se aportan dos columnas adicionales con 12 términos relacionados -en dos niveles jerárquicos diferentes- para cada término nuclear. Uno de los términos relacionados relacionados en un segundo nivel se consignó tres veces por corresponderse con ese número de aspectos diferentes. La cuarta columna refleja la cantidad de palabras clave contenedoras de los términos.

Tabla II

Principales tópicos de investigación en Ciencias Agrarias de Uruguay

Términos Nucleares	Términos Relacionados	Términos Relacionados	Palabras clave
Animales			795
	Comportamiento animal		647
		Reproductivo	166
		Productivo	98
	Enfermedades		698
		Cáncer	589
		Reproducción	673
		Indicadores-marcadores	66
		Comportamiento celular	72
		Técnicas	54
Uso del suelo			803
	Cultivos		629
		Técnicas	305
		Componentes orgánicos	398
	Pastoreo		710
		Técnicas	302

3 Conclusiones

El presente trabajo, evidenció que apoyado en las cinco funciones típicas del control de vocabulario los términos aportados incluso por público especializado requieren de la aplicación de técnicas especializadas que permitan validar la terminología del área para construir de forma más eficiente mapas disciplinarios que representen el trabajo acumulado en cualquier área del conocimiento.

Se puede decir que la investigación en Uruguay tiene dos focos principales: Los animales y el uso del suelo. El primero basado en el comportamiento reproductivo animal y el productivo (producción de leche y carne) y en las enfermedades asociadas a la reproducción y al cáncer. En menor medida, se aportan técnicas para combatirlas. El segundo foco refiere a las técnicas de cultivo y pastoreo.

Adicionalmente, puede explorarse la utilidad del control de vocabulario para la evaluación de la política en C&T de cualquier país. En el caso estudiado, si se comparan las áreas prioritarias establecidas en el PENCTI con las líneas llevadas adelante por los investigadores más prestigiosos del país se comprueba la correlación entre ambas.

Se cree conveniente el trabajo conjunto entre quienes aportan a los sistemas de información tradicionales y quienes discuten la política en C&T. Un mecanismo que permita al investigador acceder a los términos más utilizados en su área, depurados con base en las funciones del control de vocabulario podría fortalecer ambos sistemas.

Referencias bibliográficas

Aitchison, J. & Clarke.D. (2004). The thesaurus : a historical viewpoint, with a look to the future. **Cataloging & Classification Quarterly**, v. 37, n, ¾, p. 5-21.

Aitchison, J., Gilchrist, A. & Bawden, D. (2000). **Thesaurus construction and use : a practical manual**. Londres: ASLIB.

Barité, M. (2014) El control de vocabulario en la era digital: revisión conceptual. **Scire** v. 20, n.1. Disponible en: <http://ibersid.eu/ojs/index.php/scire/article/view/4196/3766> Acceso: 23 may. 2015.

Barité, M. (2015). Control de vocabulario : orígenes, evolución y proyección. **Ciência da Informação.** (Artículo aceptado para su publicación)

Barité, M., *et al.* (2013). Diccionario de Organización del Conocimiento : Clasificación, Indización, Terminología**. Montevideo: PRODIC. Disponible en:** www.prodic.edu.uy/investigacion/KOD **Acceso: 15 de may. 2015**

Bértola, L.**, *et al.*** (2005). ***Ciencia, tecnología e innovación en Uruguay: diagnóstico, prospectiva y políticas***. (Serie de Notas de Referencia) S.l: BID.

British Standards Institution (2005-2008). Structured vocabularies for information retrieval : Guide : BS 8723-1**. London: BSI. Consta de 5 partes: Parte 1: Definitions, symbols and abbreviations (2005). Parte 2: Thesauri (2005). Parte 3: Vocabularies other than thesauri (2007). Parte 4: Interoperability between vocabularies (2007). Parte 5: Formats and protocols for data exchange (2008).**

Cintra, A. M. M., *et al* (2002). Para entender as linguagens documentárias**. 2a. ed. São Paulo: Polis; APB.**

International Standard Organization (2011). International Standard ISO 25.964-1 : Information and Documentation : Thesauri and interoperability with other vocabularies. **Geneva: ISO. Consta de 2 partes: Parte 1: Thesauri for information retrieval. Parte 2: Interoperability with other vocabularies.**

National Information Standards Organization (2010). Guidelines for the construction, format and management of monolingual controlled vocabularies : an American National Standard ANSI/NISO Z39.19-2005(R2010) developed by the National Information Standards Organization**. Bethesda: NISO Press. Disponible en:** http://www.niso.org/apps/group_public/download.php/12591/z39-19-2005r2010.pdf **Acceso: 15 de may. 2015]**

Naumis, C. (2007). **Los tesauros documentales y su aplicación en la información impresa, digital y multimedia**. México: UNAM; Buenos Aires: Alfagrama.

Uruguay. Presidencia de la República (2010) Plan Estratégico Nacional en Ciencia y Tecnología. Montevideo: IMPO.

Bens culturais rurais: relato de experiência da construção de taxonomia no contexto das fazendas históricas paulistas

Mayara Cristina Bernardino
Universidade Federal de São Carlos
mayarabernardino@yahoo.com.br

Luciana de Souza Gracioso
Universidade Federal de São Carlos
luciana@ufscar.br

Maria da Graça Melo Simões
Universidade de Coimbra
gsimões@darq.uc.pt

Luzia Sigoli Fernandes Costa
Universidade Federal de São Carlos
luziasigoli@gmail.com.br

1 Introdução

Nosso escopo de pesquisa se situa no universo das fazendas históricas paulistas, responsáveis pelo cultivo de café durante os séculos XVIII e XIX no Brasil. Os modos de vida e produção construídos no entorno destas fazendas foram constitutivos da nação brasileira em diferentes aspectos. Por isto, as compreendemos como espaço de informação que demandam atenção da Ciência da informação. Nossos olhares sobre os fenômenos envolvidos com este campo vão no sentido de identificar recursos que permitam registrar, preservar e promover o acesso a estes

bens culturais e inevitavelmente a promoção do fluxo destes saberes. Alguns desses recursos passam por ações e intervenções na organização do conhecimento e, muitas das ações envolvidas com a produção de saberes neste campo, demandam por abordagens pragmáticas, que consideram a linguagem, não como representação da realidade, mas como sendo a realização de algo. Contudo estas análises teóricas não serão objeto deste trabalho, embora tenham sustentado todas as reflexões feitas ao longo do processo da pesquisa. O objetivo deste relato é descrever os resultados das pesquisas desenvolvidas, para subsidiar a elaboração de uma taxonomia a ser utilizada para a representação e recuperação de conteúdos sobre bens culturais rurais, em software criado especificamente para este fim, chamado Memória Virtual. O relato apresentado, configura-se como resultado final de pesquisa de mestrado financiada pela Fapesp e faz parte de um conjunto de resultados de pesquisas desenvolvidas em três projetos que também receberam financiamento Fapesp, coordenados por diferentes instituições (USP São Carlos – 2004/2008; UNICAMP Campinas – 2007/2012; UFSCar São Carlos – 2012-2014).

2 Fundamentos teóricos, conceituais e metodológicos

Conforme documento publicado pela Unesco, intitulado *Recomendações de Paris*, 1964, são considerados bens culturais todos os bens, sejam eles móveis ou imóveis, que possuem importância cultural para uma nação. Enquadram-se nestas categorias obras artísticas de cunho histórico, arqueológico, etnológico, espécimes (como flora e fauna) e coleções de modo geral. (UNESCO, 1964). Já o termo patrimônio, seria utilizado para designar o conjunto de bens culturais que se tornam propriedade de um país (FONSECA, 2005). Como patrimônio cultural, caracterizam-se os monumentos, conjuntos de obras, lugares notáveis (UNESCO, 1972) e como patrimônio cultural imaterial caracterizam-se as representações, os conhecimentos, instrumentos, fatos, ações ou lugares que comunidades ou indivíduos reconhecem como parte integrante de seu patrimônio cultural. Este patrimônio imaterial é retroalimentado pela própria comunidade de geração em geração. Conforme documento produzido pela Unesco (*Recomendações de Paris,* versão 2003), o patrimônio cultural imaterial se manifesta enquanto: “[...] a) tradições e expressões orais [...]; b) expressões artísticas; c) celebrações, práticas sociais, rituais e atos

festivos; d) conhecimentos e práticas, relacionados à natureza e ao universo; e) técnicas artesanais tradicionais." (UNESCO, 2003). Sobre o universo dos bens culturais, do patrimônio e também da memória, um corpo significativo de pesquisadores envolvidos com a Ciência da informação tem-se debruçado significamente nessas temáticas, especialmente no Brasil, e suas produções tem-se concentrado nos Grupos de trabalho no Enancib (Encontro Nacional de Pesquisa em Ciência da informação), a saber: GT 09 – Museu, patrimônio, informação e GT10 – Informação e memória. Foram apresentados e publicados 217 trabalhos, entre os anos de 2010 e 2013, no contexto deste evento. A urgência na discussão deste campo também se formata no tema dos trabalhos para o Enancib 2015: "Informação, Memória e Patrimônio: do documento às redes".

No âmbito destas discussões, trazemos as Fazendas Históricas Paulistas, compreendendo-as como bens culturais com perspectivas de se configurarem enquanto patrimônio cultural nacional. A partir desta linha, diferentes grupos de pesquisa têm-se voltado a alargar e aplicar metodologias para catalogar e inventariar bens culturais destas fazendas tornando-os disponíveis em um software livre chamado Memória Virtual (MV). Os projetos desenvolvidos sobre o assunto, com os quais nos relacionamos, foram:

a) "Memória Virtual de São Carlos" (2004-2008), coordenado pelo Prof. Dr. José Carlos Maldonado (ICMC/USP - Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo);

b) "Patrimônio Cultural Rural Paulista: espaço privilegiado para pesquisa, educação e turismo" (2007-2012), coordenado pelo Prof. Dr. Marcos Tognon (Unicamp - Universidade Estadual de Campinas);

c) Critérios e metodologias para a realização de inventário do Patrimônio Cultural Rural Paulista" (2012-2014), coordenado pela Profa. Dra. Luzia Sigoli Fernandes Costa (UFSCar - Universidade Federal de São Carlos).

O nosso estudo, que se integrou no projeto c), teve como objetivo geral elaborar um instrumento de linguagem, especificamente uma taxonomia, para ser utilizada como apoio à indexação de conteúdo no software MV. Subsidiaram a construção deste instrumento de representação e recuperação do conhecimento diferentes orientações teóricas e metodológicas. Entre elas dá-se especial enfoque às normas internacionais que orientam metodologicamente a sua construção como é o caso da

ANSI/NISO Z39.19-2005, que ao mesmo tempo que aponta princípios e procedimentos é também o garante internacional da sua qualidade.

O desenvolvimento da taxonomia proposta teve início em 2010, a partir da execução do projeto de pesquisa: “Organização do conhecimento no contexto das fazendas históricas do estado de São Paulo: indicações gerais para a construção de linguagens de representação da informação”, elaborada por Bernardino finalizada em 2012, enquanto pesquisa de iniciação com financiamento da Fapesp. O objetivo central de tal projeto foi o de:

a) construir bancos terminológicos relacionados a bens culturais a partir de consulta a obras científicas, literárias e técnicas relacionadas ao domínio (pautados em aporte teórico da garantia literária (BEGTHOL, 1986);
b) desenvolver entrevistas com pesquisadores do campo para identificar conceitos, categorias, relações semânticas e pragmáticas entre os conceitos relacionados ao domínio (pautados em aporte teórico da garantia do usuário especialista).

A referida pesquisa teve como resultado a construção de um banco terminológico com aproximadamente 1.000 itens, resultantes da consulta da literatura e dos especialistas.

Posteriormente, outro projeto de pesquisa foi desenvolvido, intitulado “Representação da informação de bens culturais: construindo uma taxonomia no contexto das Fazendas históricas paulistas”, desenvolvida também por Bernardino finalizada em 2015, enquanto pesquisa de mestrado feita junto ao Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade, da UFSCar, com financiamento da Fapesp. Neste projeto, o objetivo central foi o de articular os 1.000 termos já identificados, em uma estrutura taxonômica mínima, para servir de instrumento de apoio à indexação e à recuperação de conteúdos no MV. Para tanto, foram ampliadas as fontes de pesquisa para incrementar o banco terminológico e, desta vez, instrumentos de linguagem específicos como o *Thesaurus para acervos museológicos*, e o *Tesauro do Folclore e Cultura Popular Brasileira* foram utilizados.

Ao final, foram gerados 6.848 termos, sistematizados inicialmente em 16 bancos terminológicos, categorizados segundo a fonte consultada. Após análise sistemática destes conceitos, foram selecionados aqueles mais representativos para compor a taxonomia do domínio, com 3639 termos, estruturados hierarquicamente

prevendo algumas relações de equivalência, que servirá de apoio aos processos de indexação e recuperação de conteúdos no MV. As relações associativas não foram estabelecidas nesta etapa da pesquisa, por questões que tempo e cronograma, mas almeja-se me projetos futuros dar seguimento ao desenvolvimento destas relações. Para esta etapa da pesquisa, além dos aportes metodológicos já indicados nos orientamos na publicação de Simões (2008) sobre as relações conceituais em vocabulários controlados e pontualmente nas definições de taxonomia sistematizadas por Aganette, Café e Rocha (2010).

O Padrão de Descrição da Informação (PDI) e o Memória Virtual (MV): características e desafios

O Padrão de Descrição da Informação (PDI) propõe a descrição de uma variedade de tipologias de bens patrimoniais, tanto os dados gerais quantos os específicos pertinentes a um determinado tipo de bem, prevendo dessa forma, uma grande base de dados com interoperabilidade entre os diferentes tipos de bens patrimoniais inventariados, possibilitando que as informações originalmente isoladas sejam usadas de forma integrada.

A primeira versão do PDI foi inicialmente elaborada no projeto “Memória Virtual de São Carlos”, momento em que também se criou o um protótipo do software Memória Virtual. A segunda versão deste padrão foi desenvolvida no projeto de políticas públicas “Patrimônio Cultural Rural Paulista: Espaço Privilegiado para a Pesquisa, Educação e Turismo” e a fase de aplicação, testes e validação foi executada

Grupo	Conteúdo do campo	Atributos do bem patrimonial
G1	Identificação da fazenda	Dados sobre a propriedade, o proprietário e o gestor, endereço e coordenadas geográficas, o tipo de propriedade, a proteção e legislação incidente, imagens e demais informações sobre o contexto histórico-cultural em que se insere a propriedade.
G2	Informações gerais do bem patrimonial	Dados sobre a natureza e tipologia do bem, o registro, o titulo, o nível de descrição[8] (arquitetônico, arquivístico, bibliográfico, bem móvel integrado, iconográfico e natural), localização física com as devidas coordenados do bem.
G3	Autoria	Dados relacionados com a variedade de tipos de autoria ou responsabilidades atribuídas à criação do bem descrito.
G3	Cadastro de autoria	Dados complementares ao Grupo 2 como nome, pseudônimo, função ou atividade e período de existência.
G5	Produção	Dados sobre o local, ano, versões, edições e responsabilidades concernentes a produção do bem patrimonial.
G6	Descrição	Dados de conteúdo, estilo, resumo, características físicas, dimensões e volumetria, análise estilística, iconográfica e o contexto e espacial e temporal do bem descrito[9] e meio de acesso.
G7	Audio-Visual	Dados e materiais produzidos pela pesquisa realizada durante o inventário.
G8	Diagnóstico e Intervenção	Dados sobre o estado de conservação do bem, as intervenções realizadas com os seus respectivos responsáveis.
G9	Disponibilidade, uso e proteção.	Dados sobre a situação de disponibilidade, condição de acesso, uso, reprodução, aproveitamento e proteção do bem patrimonial descrito.
G10	Histórico da procedência	Dados sobre os tipos de transação, valor venal, data de aquisição do bem patrimonial.
G11	Assunto e descritores	Palavras-chave. Termo livre da linguagem natural e termos controlados, retirados de um vocabulário pré-definido, que traduzem o conteúdo temático do bem patrimonial10.
G12	Fonte de informação	Dados sobre as fontes de informação utilizadas na pesquisa histórica do bem patrimonial, independente da sua natureza e tipo de bem patrimonial.
G13	Responsável pela pesquisa	Nome do responsável pela pesquisa histórica, data da coleta de dados e anotações complementares feitas pelo pesquisador.

durante o projeto “Critérios e metodologias para a realização de inventário do Patrimônio Cultural Rural Paulista”, finalizado em 2014. Na figura 1, são indicados os 13 campos iniciais gerados para descrição de dados, que se desdobram em 40 subdivisões no PDI.

Figura 1 - Campos do Padrão de Descrição da informação.
Fonte: **Nakagawa, Costa e Scarpeline (2010, p.48). (Onde se lê G3, deve-se considerar como G4).**

Sobre o grupo 11 (Assuntos e Descritores), nos debruçamos ao longo desta pesquisa no intuito de diagnosticar e sugerir o desenvolvimento de um instrumento de linguagem que servisse de apoio à indexação dos bens que serão cadastrados no MV. Apesar deste instrumento ter sido desenvolvido, as políticas relativas ao seu uso e alimentação ainda precisam de ser definidas. Interessa esclarecer que a denominação dos campos “Assunto” e “Descritores” foi estabelecida pela equipe interdisciplinar desenvolvedora do Padrão de Descrição da Informação. O seu desenvolvimento teve em atenção o fato de que os usuários/indexadores no sistema poderem não ser necessariamente, bibliotecários ou especialistas em documentação e arquivo. Potencialmente o sistema contará com cadastros feitos por proprietários e funcionários das próprias fazendas. Por este contexto é que se optou em utilizar a categoria “Assunto” para registro de palavras livres, linguagem natural, assuntos gerais e “Descritores” para assuntos específicos e controlados a partir do uso da Taxonomia proposta.

O PDI fundamentou a concepção da estrutura do Memória Virtual (MV). Software livre desenvolvido por pesquisadores do curso de Ciência da Computação da Universidade de São Paulo (ICMC/USP), no *campus* de São Carlos – SP, com a coordenação e supervisão da Profa. Dra. Elisa Yumi Nakagawa, que parceiros dos projetos já citados. O MV esta finalizado e em fase de testes. A taxonomia

desenvolvida será incorporada como um banco de dados independente ao sistema Memória Virtual para ser consultada. A interface inicial para cadastro de bens culturais pode ser conferida na Figura 2:

Figura 2 - Interface para cadastro de bem patrimonial no Memória Virtual
Fonte: *Website* do Memória Virtual Disponível em: http://143.107.231.114:8080/memoriavirtual/login.jsf

Cabe também acrescentar que outras pesquisas voltadas ao desenvolvimento de políticas de indexação, uso e avaliação da taxonomia no MV estão sendo elaboradas no âmbito do Grupo de pesquisa “Patrimônio cultural: memória, preservação e gestão sustentável” credenciado no CNPQ. Ainda, há um conjunto de pesquisas e investigações sendo feitas por pesquisadores deste grupo (GRACIOSO, 2010), voltadas a discussão e análise da linguagem, em sua perspectiva pragmática, isto é, de uso, que intentam analisar o movimento de significação dos processos de comunicação, refletindo os desdobramentos que tal abordagem poderia sugerir na configuração de instrumentos de representação da informação em contextos de memória e de bens culturais.

Estruturação inicial da taxonomia sobre bens culturais para uso no Memória Virtual

A estrutura metodológica inicial para organização terminológica se pautou na estruturação geral de taxonomias, tal qual sugere a Norma ANSI/NISO Z39.19-2005, mas com incorporações de alguns aspectos metodológicos adaptados as especificidades do domínio. Inicialmente a construção da taxonomia teve a seguinte orientação:

1 Elaboração da taxonomia:

a) Levantamento de termos pautados em garantia literária e do especialista;

b) Organização e sistematização dos conceitos identificados e selecionados em bancos de conceitos – ordenados alfabeticamente;

c) Seleção de categorias gerais para sistematização conceitual;

d) Inclusão dos termos em categorias conceituais sugeridas;

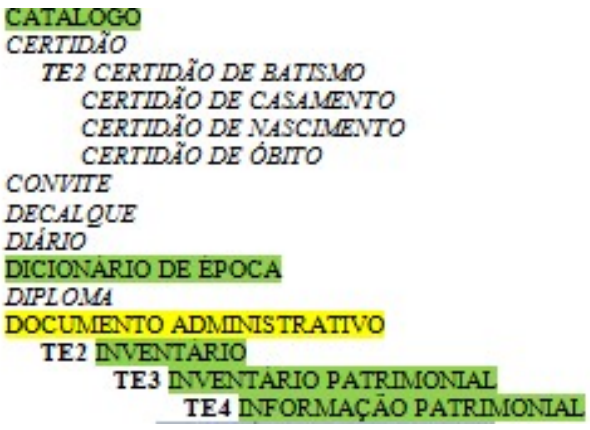
e) Estabelecimento de relações de hierarquia e de sinonímia entre os termos dentro das categorias;

f) Delimitação de termos preferidos para uso na indexação.

Baseando-se nos métodos de construção de taxonomias, apresentados por González (2011), em que se pode fazer a reutilização de estruturas já prontas complementando-as, a fim de economizar tempo e esforços, optou-se por utilizar a estrutura hierárquica desenvolvida por Ferrez e Bianchini (*Thesaurus para acervos museológicos*, 1987) e alimentá-la com os termos elencados e sistematizados nos bancos terminológicos criados sobre o domínio das fazendas históricas.

Um recurso metodológico utilizado, porém, não indicado em orientações normativas, foi o uso de cores na apresentação final do instrumento, que indicam o banco terminológico de origem do termo. Cada um dos bancos terminológicos recebeu uma cor para seu armazenamento, ao passo em que os termos oriundos destes bancos foram incorporados na taxonomia, estas cores foram mantidas no intuito de permitir uma visualização gráfica sobre a origem e as garantias envolvidas no contexto do termo. O resultado pode ser verificado na figura 3, em que se apresenta um extrato da taxonomia desenvolvida.

Figura 3: Exemplo de estruturação da terminologia empregando cores para sinalizar a origem dos bancos terminológicos utilizados. (Verde: banco terminológico de entrevistas com pesquisadores; Amarelo: banco terminológico do Relatório Parcial FAPESP.Projeto de Políticas Públicas Patrimônio Cultural Rural Paulista: Espaço Privilegiado para Pesquisa, Educação e Turismo. Os termos em itálico, sem cores, são oriundos do *Thesaurus para acervos museológicos*.



Fonte: As autoras.

Toda a estrutura da taxonomia proposta, assim como detalhes sobre a configuração dos bancos terminológicos estão disponíveis como apêndice no trabalho: BERNARDINO, M. C. *Representação da informação de bens culturais: construindo uma taxonomia no contexto das Fazendas históricas paulistas*. São

Carlos: UFSCar, 2015. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade). Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade (PPGCTS), Universidade federal de São Carlos, 2015.

Considerações finais

Com o intuito de registrar, sistematizar, preservar e promover o acesso aos bens culturais oriundos das fazendas históricas de café do interior do estado de São Paulo, no Brasil, foram desenvolvidos, ao longo de 10 anos, três grandes projetos de pesquisa, financiados pela FAPESP, envolvendo equipe interdisciplinar e interinstitucional. Os principais resultados destas pesquisas, foram o desenvolvimento de um Padrão de Descrição da Informação (PDI) utilizado em um software livre que foi desenvolvido no âmbito destes projetos, intitulado Memória Virtual (MV). Com o propósito de promover, no âmbito da organização do conhecimento, um instrumento de linguagem que pudesse colaborar com a indexação e recuperação de conteúdos neste sistema, foram desenvolvidos dois projetos de pesquisa (de graduação e mestrado), ambos com financiamento da Fapesp, que tiveram como objetivo construir bancos terminológicos relacionados ao domínio das fazendas históricas e arranjo destes termos em uma estrutura taxonômica. Como resultado destes projetos, foi desenvolvida uma taxonomia sobre bens culturais rurais, com 3.639 termos, organizados sistematicamente, que será vinculada ao Memória Virtual. Esta linguagem será utilizada por uma variedade indefinida de sujeitos, que poderão participar de atividades de registros de bens patrimoniais neste sistema e também poderão consultá-lo.

Este elemento, somado a natureza dos bens culturais, tornam a construção deste instrumento de linguagem desafiador e conflitante quanto aos seus propósitos. Mesmo assim, acreditamos que o exercício desenvolvido ao longo desta pesquisa possa servir ao campo da organização do conhecimento, talvez menos pelo seu resultado instrumental, mas mais pelas problematizações que se concretizaram no que diz respeito ao desafio de restringir a linguagem a sua função de representação. Diante disto nos convencemos de que os espaços de memória (mas não só eles) demandam posturas pragmáticas de análise, que privilegiem as ações sociais vivenciadas e construídas por intermédio dos diferentes usos da linguagem. A imagem que encerra este artigo nos faz refletir sobre os limites da nossa linguagem para

representarmos as memórias vividas, e nos aguça a querer continuar na busca da compreensão sobre a linguagem enquanto um instrumento de ação.

Figura 4: Senzala – Fazenda Santa Maria do Monjolinho, São Carlos, SP

Fonte: Os autores, 2012.

Referências

AGANETTE, Elisangela; ALVARENGA, Lídia; SOUZA, Renato Rocha. Elementos constitutivos do conceito de Taxonomia. **Inf. & Soc.:** Est., João Pessoa, v. 20, n. 3, p. 77-93, set./dez. 2010. Disponível em: <http://www.ies.ufpb.br/ojs2/index.php/ies/article/view/3994>. Acesso em: 15 de maio 2012.

ANSI/NISO Z39.19-2005*:* guidelines for the construction, format, and management of monolingual controlled vocabularies. Bethesda: NISO, 2005. 172 p.

BERNARDINO, M. C. **Organização do conhecimento do contexto das fazendas históricas do estado de São Paulo:** indicações gerais para a construção de linguagens de representação da informação. Trabalho de Conclusão de Curso (graduação). Universidade Federal de São Carlos. São Carlos, 2012.

BERNARDINO, M. C. **Representação da informação de bens culturais: construindo uma taxonomia no contexto das Fazendas históricas paulistas.** São Carlos: UFSCar, 2015. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade). Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade (PPGCTS), Universidade federal de São Carlos, 2015.

COELHO, Marisa Colmago. **Tesauro do Folclore e Cultura Popular Brasileira.** Disponível em http://www.cnfcp.gov.br/tesauro/sistematica.html Acesso em set. 2014.

FERREZ, H. D.; BIANCHINI, M. H. S. **Thesaurus para acervos museológicos**. Fundação Nacional Pró-Memória, Rio de Janeiro, v. 1, 1987. 86p. (série técnica).

GONZÁLEZ, J. A. M. **Linguagens documentárias e vocabulários semânticos para a web:** elementos conceituais. Salvador: EDUFBA, 2011. 128 p.

GRACIOSO, Luciana de Souza. Parâmetros teóricos para elaboração de instrumentos pragmáticos de representação e organização da informação na Web: considerações preliminares sobre uma possível proposta metodológica. **InCID: Revista de Ciência da Informação e Documentação**, Brasil, v. 1, n. 1, p. 138-158, jul. 2010. ISSN 2178-2075. Disponível em: <http://www.revistas.usp.br/incid/article/view/42310>. Acesso em: 24 Mai. 2015. doi:http://dx.doi.org/10.11606/issn.2178-2075.v1i1p138-158.

HJØRLAND, B. Domain analysis in Information Science: Eleven approaches – traditional well as innovative. **Journal of Documentation**, v.58, n.4, p.422-462, 2002.

HJØRLAND, B. The classification of Psychology: A case study in the classification of a knowledge field. **Knowledge Organization**, v.25, n.4, p.162-201, 1998.

NAKAGAWA, E. Y.; COSTA, L. S. F.; SCARPINELI, R. **Fazendas históricas paulistas do século XVIII e XIX: premissas teóricas e metodológicas para inventariar bens patrimoniais.** *Resgate*, Campinas, v. 18, n. 20, 2010, p. 37-53. Disponível em: <http://www.cmu.unicamp.br/seer/index.php/resgate/article/view/305/301>. Acesso em: 22 out. 2013.

SIMÕES, M. G. M. **Da abstracção à complexidade formal**: relações conceptuais num tesauro. Coimbra: Almedina, 2008.

UNESCO, **Recomendação de Paris**, 1964. Recomendação sobre medidas destinadas a proibir e impedir a exportação, a importação e a transferência de propriedade ilícitas de bens culturais. Disponível em: <http://www.iphan.gov.br/baixaFcdAnexo.do?id=4676> Acesso: jan. 2015

UNESCO, **Recomendação de Paris**, 1972. Convenção sobre a Proteção do Patrimônio Mundial, Cultural e Natural. Disponível em :<http://portal.iphan.gov.br/portal/baixaFcdAnexo.do?id=244> Acesso: jan. 2015

UNESCO, **Recomendação de Paris**, 2003. Convenção para a salvaguarda do patrimônio cultural imaterial. Conferência geral da Organização das Nações Unidas para a Educação, a Ciência e a Cultura doravante denominada “UNESCO” em sua 32ª sessão. Disponível em <http://portal.iphan.gov.br/portal/baixaFcdAnexo.do?id=271> Acesso jan. 2015.

Harmonization of CIDOC CRM ontology in the context of archives, libraries and museums

Laís Barbudo Carrasco
Universidade Estadual Paulista
laiscarrasco@hotmail.com

Silvana Ap. Borsetti Gregório Vidotti
Universidade Estadual Paulista
vidotti@marilia.unesp.br

Phil Manfred Thaller
Universität zu Köln
manfred.thaller@uni-koeln.de

Introduction

By the fact that Digital Cultural Heritage Repositories, such as Libraries, Archives and Museums, use different metadata standards to describe their information resources, the metadata harmonization from the cultural heritage field is a challenge, because the data models are more designed on the community requirements than on requirements of cross-community interoperability.

In order to integrate information from heterogeneous sources, ontologies as semantic technologies, are already being used and CIDOC Conceptual Reference Model (CRM) is a very prominent ontology used for such purposes.

> The **CIDOC CRM** is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be

> mapped to. […] In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. (CIDOC CRM)

Metadata are forms used to describe, manage, catalog, and classify documents (information objects). In Libraries, they are traditionally used to describe books, in other words, bibliographic data. In addition, in museums, metadata describes artifacts, paintings, and sculptures. On other hand, in Archive, the metadata are used to describe finding aids, charters, and official documents. Therewith, this treated information can be effectively retrieval. The cultural memory institutions aim to treat their information in order to obtain access and retrieval.

A way to make metadata interoperate is mapping them into an ontology. CIDOC Conceptual Reference Model (CIDOC CRM) is an ontology from the cultural heritage domain, which has been created as a tool for information integration.

Cultural heritage institutions such as museums, archives, or libraries are confronting a crescent necessity to integrate their system. For this reason, to make cultural resources accessible, it is necessary the use of rich metadata structures, capable to cover the variety of material held in these memory institutions.

In this context, ontologies are used as an important tool for achieving information integration, in other words, metadata can be semantically mapped and integrated into an ontology, which has the competence not only to conceptualize specific domains, but also to express their semantics.

The CIDOC CRM arised from the CIDOC Documentation Standards Group in the International Committee for Documentation of the International Council of Museums and CIDOC CRM was accepted as the ISO 21127 in 2006. The purpose of the CRM model is to provide a common language for heterogeneous information systems, and to permit their integration, despite possible semantic and structural incompatibilities. In that way, cultural heritage information can be exchanged and retrieved; and cultural heritage institutions can make their information systems interoperable without having to compromise their specific needs or the current level of precision of their data. (CIDOC CRM)

In summary, CIDOC CRM is a high-level ontology to permit information integration for cultural heritage data and their correlation with museum, library and archive information, which is easily converted to other machine readable formats such

as RDF and XML. Possibly the most ambitious application of the CRM is in the development of integrated query tools, mediation systems and data warehouses. At present, much of the information stored in library catalogues, archival finding aids and museum collection management systems remains isolated. Different information resources normally need to be queried individually, and cross-system links are rare. The ability to combine and integrate information from multiple sources has the potential to add significant value to existing data - facilitating research and enhancing the quality of the user's experience. (ARTUR; CROFTS; LE BOEUF; 2002).

When a user makes a search in a data base he or she doesn't look for a library resource or an archival resource or a museum resource, the user searches for information. In this way, when the user insert a query into the search engine the user does not take into consideration the nature of the resource, that's why it's very important to develop a harmonization of CIDOC CRM ontology in the context of archives, libraries and museums, because it will be able to understand how the CRM entities can be interpreted in the context of archives, libraries and museums and then it will be possible to potentialize the information retrieval.

State of art

Library, Archive, and Museums are repositories of collective memory and they have some functions, for example, the maintenance and conservation of collections; the exposure; and provision of the means for the objects or information retrieval.

Storing, cataloging, classifying and exhibiting objects are part of the relationship process of the man with the world, and they are operations that do not arise with the museum, nor with library, and nor with the archive, but with collections.

Library, Archive and Museum – all the three institutions preserve the memory and organize the access to information. On other hand, according to the institution, the role assigned to the document is different. Traditionally, libraries saves book, archives saves legal documents and museum masterpieces. Thus, each typological material has a way to be treated. In this way, it's needed to differentiate what a term refers to each institution. According to the author Alvares (s.d.) archive, library and museum have the following meanings:

- Archive: Ordained storage of documents, created by a person or institution in the course of its activity, and preserved to achieve their goals, aiming the utility that

they may offer in the future, therefore, the archival information is information accumulated by institution to prove or inform, as the institution pursues its purposes/ends, with whom it relates, which rights and duties it represents.

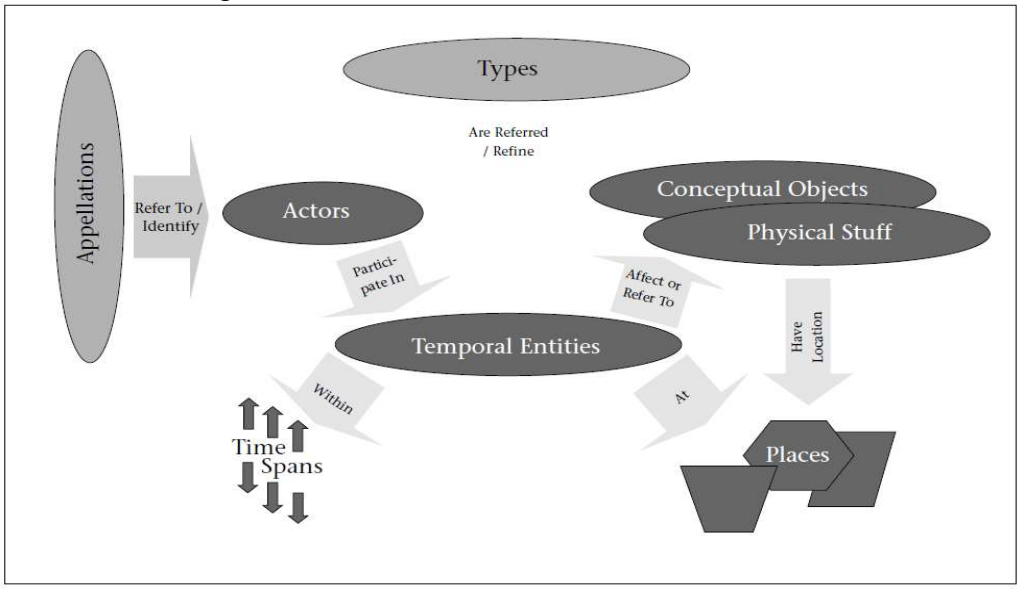
- Library: Collection of books, periodicals, audiovisual documents, among others, selected based on their usefulness and arranged to provide access to materials and information, therefore, the organization of the documents in the library reflects the organization given to the area of knowledge (or the subject specifically covered by the documentation center).

- Museum: It is a permanent institution, nonprofit organization, in service of society and of its development, open to the public which acquires, conserves, researches, disseminates and exposes the material evidence of people and their environment, for education and delight of society; therefore, the museum, as the instance of the memory representation and social space in which information is considered cultural input, is a fertile ground for the development of studies and actions related to information.

Result

The CRM scope can be defined as all the necessary information for the scientific documentation of cultural heritage collections, in order to enable a broad exchange of information from the area and the integration of heterogeneous sources.

Figure 1 – The main model entities.



Source: Cidoc CRM (DOERR, 2003).

CIDOC CRM contains classes and logical groups of properties. These groups have to do with the notions of participation, structure, location, evaluation and identification, purpose, motivation, usage, and so on. These properties have put temporal entities and with them, the events in a central location (LIMA, 2008).

Table 1 – CIDOC CRM Harmonization in the context of Libraries, Archives and Museums.

CIDOC CRM Entity	**Library**	**Archive**	**Museum**
E1 CRM Entity	CRM Entity about **bibliographic material.**	CRM Entity about **archival material.**	CRM Entity about **museological material.**
E18 Physical Stuff	Literature ("Book") about the Physical Stuff.	Documentation ("Charter") about the Physical Stuff.	Museum object
E28 Conceptual Object	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of bibliographic material.	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of archival material.	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of museum material.
E39 Actor	Author.	Institution.	Artist.
E41 Appellation	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context refered to the Library and the bibliographic material.	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context refered to the Archive and the archival material.	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context refered to the Museum and museological material.
E50 Date	Date of publication.	Date of fund.	Date of creation.
E52 Time-Span	Time-Span of the "Book, Collection".	Time-Span of the "Charter, Fund".	Time-Span of the "Object, Collection".

E53 Place	Place of publication or storage (Library).	Place of provenance or storage (Archive).	Place of creation or storage (Museum).
E55 Type	Typological distinctions of "Book, Collection".	Typological distinctions of "Charter, Fund".	Typological distinctions of "Object, Collection".

Source: Authors.

In this section, it was built from the CIDOC CRM vocabulary a correspondence of terms handled into the cultural heritage institutions – Library, Archive and Museum; because CIDOC CRM entities can have different meanings facing the diverse types of material handled in the three sisters institutions.

Final considerations

We are facing an explosion of varieties of cultural heritage metadata schemas. Cultural Heritage - as the safeguard of the memory - has heterogeneous materials, and for this reason they are safeguarded, protected and conserved in the Libraries, Archives and Museum. In order to treat these digital objects, metadata standards are used. On other hand, how can these three institutions exchange data?

According to the literature, there are many XML metadata mapping to the CIDOC CRM ontology efforts, since this ontology is considered one of the most appropriate models in integration architectures.

In this way, metadata can be mapped into an ontology to provide interoperability of its data and also to achieve information integration. When the different kind of metadata are mapped into an ontology the system can interoperate and the information access is higher as well as their information retrieval.

“An ontology is a specification of a conceptualization” (GRUBER, 1993). More specially, the CIDOC CRM ontology is the specification of the Cultural Heritage conceptualization. CIDOC CRM has an abstract hierarchy.

In summary, the metadata from the three sisters institutions – Library, Archive and Museum – can be mapped into the ontology CIDOC CRM to provide interoperability and information integration of these heterogeneous data from the digital cultural heritage systems.

References

ALVARES, Lillian. (s.d.) **Informação em Arquivos, Bibliotecas e Museus**. (Slides) - Faculdade de Ciência da Informação - Universidade de Brasília. Available at: http://www.alvarestech.com/lillian/Analise/Modulo1/Aula13ABMII.pdf. Access: May, 2012.

ARTUR, O.; CROFTS, N.; LE BOEUF, P. Elag presentation ontologies. Semantic web and libraries. In: LIBRARY SYSTEMS SEMINAR, 26., 2002, Rome. **Proceedings**... Rome: European Library Automation Group, 2002. Disponível em: <http://www.ifnet.it/elag2002/papers/pap9.html>. Acesso em: 18 abr. 2012.

CIDOC CRM. Disponível em:<http://www.cidoc-crm.org/>. Acesso em: 20 fev. 2014.

CROFTS et al. (Ed.). **Definition of the Cidoc Conceptual Reference Model**. Cidoc, Icom, 2005. Disponível em: <http://www.cidoc-crm.org/docs/cidoc_crm_version_4.2.pdf >. Acesso em: 28 jan. 2015

CROFTS, N. **Museum informatics:** the challenge of integration. 2004. Tese (Doutorado em Ciências Econômicas e Sociais) – Faculté des Sciences Économiques et socials, Universidade de Genebra, Genebra, 2004.

DOERR M., The Cidoc CRM: an ontological approach to semantic interoperability of metadata. **AI Magazine**, v. 24, n. 3 p. 75-92, 2003.

DOERR, M.; STEAD, S. **The Cidoc CRM:** standard for the integration of cultural information. Glasgow, Jan. 29, 2008. Disponível em: <http://www.uniurb.it/sbc/ist_bal/seminario/carlo%20meghini/CIDOC%20CRM%20introduction.pdf>. Acesso em: 20 maio 2012.

DOERR et al. The Europeana Data Model (EDM). In: IFLA GENERAL CONFERENCE AND ASSEMBLY, 76., 2010, Gothenburg. **Proceedings**… Gothenburg: Ifla, 2010. Disponível em: <http://conference.ifla.org/past-wlic/2010/149-doerr-en.pdf>. Acesso em: 18 fev. 2012.

GRUBER, T. R. **What is an ontology?** Stanford: Stanford University, 2001. Disponível em: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>. Acesso em: 27 fev. 2012.

GRUBER, T. R. Towards principles of the design of ontologies used for knowledge sharing. In: INTERNATIONAL JOURNAL HUMAN-COMPUTER STUDIES, SUBSTANTIAL REVISION OF PAPER PRESENTED AT THE INTERNATIONAL WORKSHOP ON FORMAL ONTOLOGY, Padova, 1993. **Proceedings**… Padova: Standford University, 1993. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.89.5775>. Acesso em: 20 jan. 2014.

LE BOEUF, P., Using an ontology-driven system to integrate museum information and library information. In: SYMPOSIUM ON DIGITAL SEMANTIC CONTENT ACROSS CULTURES, 2006, Paris. **Proceedings**… Paris: Louvre, 2006.

LIMA, J. A. O. ***Modelo genérico de relacionamentos na organização da informação legislativa e jurídica.*** 2008. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, DF, 2008.

Catalogación de recursos interactivos: análisis de recomendaciones y prácticas en catálogos internacionales

Víctor Amor
Investigador independiente
victoramorss@gmail.com

Daniel Martínez-Ávila
Universidade Estadual Paulista
dmartinezavila@marilia.unesp.br

Rosa San Segundo
Universidad Carlos III de Madrid
rsan@bib.uc3m.es

1 Introducción

La organización y representación de recursos interactivos se trata de un creciente desafío para el acceso a la información en bibliotecas y aplicabilidad de estándares tradicionales en nuevos formatos. Mientras que los videojuegos son cada vez más uno de los tipos de recursos más demandados en bibliotecas tanto para su uso recreativo como educacional (Sanford 2008; Levine 2009), la preservación y descripción de estos recursos se está convirtiendo en uno de los principales problemas para uso a largo plazo y construcción de memoria. En este sentido, diversos autores han señalado la importancia y retos de la preservación de videojuegos en bibliotecas (McDonough et al. 2010; McDonough 2011; Winget 2011; Watson 2012). Tal como indica Dan Pinchbeck, de la Universidad de Portsmouth, “es

necesario catalogar y almacenar una enorme cantidad de información. En particular, los juegos no se archivan habitualmente, ya que se consideran artefactos culturales desechables de escaso valor, si bien representan una parte realmente importante de nuestra historia cultural reciente. Los juegos son uno de los formatos de medios digitales más extendidos en nuestro planeta y debemos conservarlos para generaciones futuras" (SINC/AG 2009). Henry Lowood, de la Universidad de Stanford (citado en Owens 2013), también afirma que la gestión a largo plazo de esta colección pasa por la correcta gestión de los metadatos, ontologías, terminología, etc., es decir, aspectos relacionados con la organización del conocimiento.

2 Metodología

En este trabajo se revisan los principales trabajos, guías y recomendaciones en catalogación de videojuegos y se analizan algunos de los principales catálogos internacionales de España y Estados Unidos para determinar las prácticas y grado de adopción de estas recomendaciones en catalogación. El criterio para la elección de los cinco casos fue el prestigio de las colecciones de recursos interactivos, siendo para el caso de España los catálogos de la Biblioteca Nacional y de la Comunidad de Madrid, dos de los más exhaustivos, y en el de los Estados Unidos, además del de la Library of Congress, los de la University of Oregon y University of North Carolina, dos de los más completos respecto a este tipo de recursos.

3 Guías y recomendaciones de catalogación de videojuegos

Diversos trabajos han tratado la temática de la organización y representación de videojuegos en bibliotecas. James Paul Gee (2007), por ejemplo, ofrece pautas para la inclusión de juegos en bibliotecas en relación a la educación. Más recientemente, Jerome McDonough y otros (2010a; 2011) han discutido la posibilidad de aplicar el modelo conceptual FRBR (Functional Requirements for Bibliographic Records, desarrollado por la IFLA en 2009) a los videojuegos para determinar los campos e información necesaria para recuperar este tipo de recursos en los catálogos. Algunos recursos que pueden proporcionar la información necesaria para la catalogación incluyen la "International Game Developers Association"

(http://www.igda.org/), que proporciona orientación para la catalogación de los distintos responsables que aparecen en los créditos de la obra, y el sitio “MobyGames. The Authoritative Video Game Database” (http://www.mobygames.com/) que también proporciona estos datos.

Como iniciativa más notable de buenas prácticas en el ámbito bibliotecario, cabe destacar el proyecto de la American Library Association (ALA) Team G (2011), en el cual se ofrece información sobre aspectos como el desarrollo de una colección de videojuegos, circulación y acceso, selección y compra, expurgo y, como aspecto más importante para los objetivos del presente estudio, el registro MARC ideal. La información de este proyecto fue recogida tras contactar con bibliotecas públicas, escolares y universitarias de los Estados Unidos y Canadá y expertos en el área.

En el informe final se indica que como mínimo un registro debe tener información sobre el título del videojuego, el sistema al que el juego pertenece, el año y el editor que lo publica, la clasificación por edades, el número jugadores posibles, si tiene o no multijugador en línea, sinopsis del juego, si es necesario utilizar algún periférico específico (por ejemplo Kinect, Wii Motion Plus), descripción del producto, y el precio. Se indica que el estudio desarrollador es normalmente el “autor” de la obra. Además del estudio, otras menciones de responsabilidad que se deben incluir son director, director de arte, director del programa, y actores de voz. En los casos de bibliotecas que incluyen también consolas en préstamo, la entrada en el catálogo debería tener la siguiente información: el sistema, ítems que deben ser prestados junto a la consola (como mandos, cables, etc.), y precio.

Finalmente, la ALA también ofrece como posibilidades para los catalogadores utilizar otros estándares como Categories for the Description of Works of Art (CDWA), la cual es utilizada principalmente con juegos descargables y online, y el esquema general Dublin Core.

4 Análisis de registros MARC en catálogos internacionales

Como primer ejemplo se analizó la catalogación y clasificación del juego “Age of Empires”, por las bibliotecas de la Comunidad de Madrid, España (ver Figura 1).

En el campo 245 (“Title Statement”) se encabeza por el título. En el campo 250 (“Edition Statement”) se podría hacer mención de edición, pero en el caso del ejemplo (que es una edición de coleccionista) se hace constar también en el campo 245. El

campo 260 (“Publication, Distribution, etc. (Imprint)”) pone la última fecha en que se ha publicado/distribuido y consta el editor (Microsoft) pero no la empresa desarrolladora (Ensemble Studios). El campo 300 (“Physical Description”) hace constar el soporte del recurso interactivo, sea este CD-ROM u otros soportes. En el campo 300 no constarían los recursos web, estos podrían estar en el campo 856 (“Electronic Location and Access”). Los campos 336 y 337 son dignos de mención; el campo 336 (“Content Type”) en MARC ayuda a especificar el tipo de contenido e incluye “Programa”; el campo 337 (“Media Type”) incluye “electrónico”. El campo 538 (“System Details Note”) como campo de notas de datos técnicos sería útil a la hora de catalogar este tipo de recursos, pero no es utilizado. En el caso del ejemplo, el campo 500 (“General Note”) incluye las características del equipo informático que permite su uso. El campo 594 (“Local Notes”) se podría utilizar para incluir a la empresa desarrolladora, distinta del editor, pero tampoco es utilizado. El campo 753 (“System Details Access to Computer Files”) se podría usar con estos recursos interactivos, utilizando $a para la marca y modelo de la máquina, $b para el sistema operativo en el caso de juegos de ordenador o teléfonos móviles con un sistema operativo propio, etc. pero tampoco es utilizado

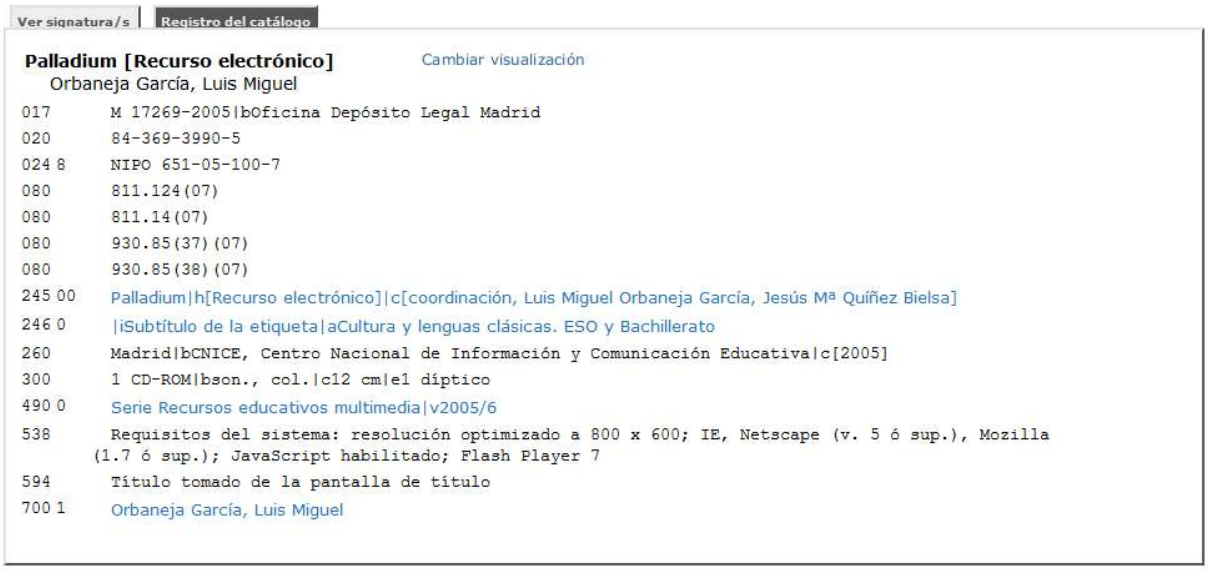
El campo 650 (“Subject Added Entry – Topical Term”) no suele ser muy utilizado en recursos interactivos, en el caso del ejemplo tan solo muestra “Juegos de ordenador”. El campo 655 (“Index Term – Genre/Form”), término de indización con el subcampo $a como Género/forma, sería una manera de catalogación más profunda de este medio, pero tampoco es utilizado. El valor que muestra la Comunidad de Madrid en el campo 080 (“Universal Decimal Classification Number”) es el 793.7 (086) en la categoría juegos, tal vez convendría también el número 621.397.48 Equipos de video y televisión anejos (juegos, teletexto, etc.) dado el uso tecnológico de los mismos en previsión de una clasificación más estructurada y definida a posteriori en el universo de la información de estos recursos interactivos.

Figura 1: campos MARC y situación en la CDU de los recursos interactivos de las bibliotecas de la Comunidad de Madrid

DOCUMENTO
EJEMPLARES (5)
COMENTARIOS
ETIQUETAS
Visualización MARC
001 ES-MaRBCM00217257c-2
003 ES-MaRBCM
005 20140826:05085200
008 120620s2000 sp r b spa d
040 $aM-BPCM $bspa $cM-BPCM
080 $a793.7(086)
090 $aBPCM -- $jCDR 793 AGE
245 00 $aAge of Empires [$hRecurso electrónico] $b: edición para coleccionistas
260 $a[USA] : $bMicrosoft, $c2000
300 $a4 CD-ROM + $e2 manuales + 2 h. pleg.
336 $aPrograma $2isbdcontent
337 $aelectrónico $2isbdmedia
500 $aCaracterísticas del sistema: Windows 95/98/NT 4.0 o superior ; Pentium II 90 Mhz ; 16 ó 32 Mb de RAM (según el juego) ; 100, 200 ó 235 Mb de espacio disponible para la instalación típica (según el juego) ; Lector CD-ROM 4 x o superior ; Super VGA (con 1 Mb de VRAM) ; Tarjeta de sonido compatible con la API ; Módem de 28,8 Kbps para jugar en conexión Internet
505 0 $aDisco [1.] Age of Empires : edición gold -- Disco [2.] Age of Empires II : the age of kings -- Disco [3.] Age of Empires II : the conquerors : expansion -- Disco [4.] Age of Empires : banda sonora
650 4 $aJuegos de ordenador
650 4 $aMúsica para cine

Como segundo ejemplo se analiza el caso de la Biblioteca Nacional de España (BNE) (ver Figura 2). La forma de catalogación en la BNE difiere de la Comunidad de Madrid en que utilizada el campo 538 (System Deatails") y no el 500 ("General Note") para los requerimientos técnicos necesarios para utilizar el recurso. Curiosamente, en esta nota no se explicita el sistema al que pertenece. Los campos de la CDU (080) hacen referencia a la historia, latín y griego, relacionados con el contenido del recurso (aboutness) y no con el formato.

Figura 2: registro MARC de recurso electrónico en la Biblioteca Nacional de España



Ver signatura/s | Registro del catálogo

Palladium [Recurso electrónico] Cambiar visualización
Orbaneja García, Luis Miguel

```
017      M 17269-2005|bOficina Depósito Legal Madrid
020      84-369-3990-5
024 8    NIPO 651-05-100-7
080      811.124(07)
080      811.14(07)
080      930.85(37)(07)
080      930.85(38)(07)
245 00   Palladium|h[Recurso electrónico]|c[coordinación, Luis Miguel Orbaneja García, Jesús Mª Quíñez Bielsa]
246 0    |iSubtítulo de la etiqueta|aCultura y lenguas clásicas. ESO y Bachillerato
260      Madrid|bCNICE, Centro Nacional de Información y Comunicación Educativa|c[2005]
300      1 CD-ROM|bson., col.|c12 cm|e1 díptico
490 0    Serie Recursos educativos multimedia|v2005/6
538      Requisitos del sistema: resolución optimizado a 800 x 600; IE, Netscape (v. 5 ó sup.), Mozilla
         (1.7 ó sup.); JavaScript habilitado; Flash Player 7
594      Título tomado de la pantalla de título
700 1    Orbaneja García, Luis Miguel
```

En la Library of Congress de los Estados Unidos existe una guía para codificar recursos electrónicos (Library of Congress, 2010). Esta guía se usa para software de ordenador (incluyendo programas y juegos) y para "ordenadores orientados a lo multimedia", sistemas o servicios online. El ejemplo utilizado en la LOC para la catalogación de juegos de ordenador es el siguiente (Figura 3):

Figura 3: Ejemplo de juego de ordenador en la guía de la LOC

Computer game

The resource is a game that allows multi-user play online. The type of file is *game*.

Example(s):

World of warcraft, a computer game.

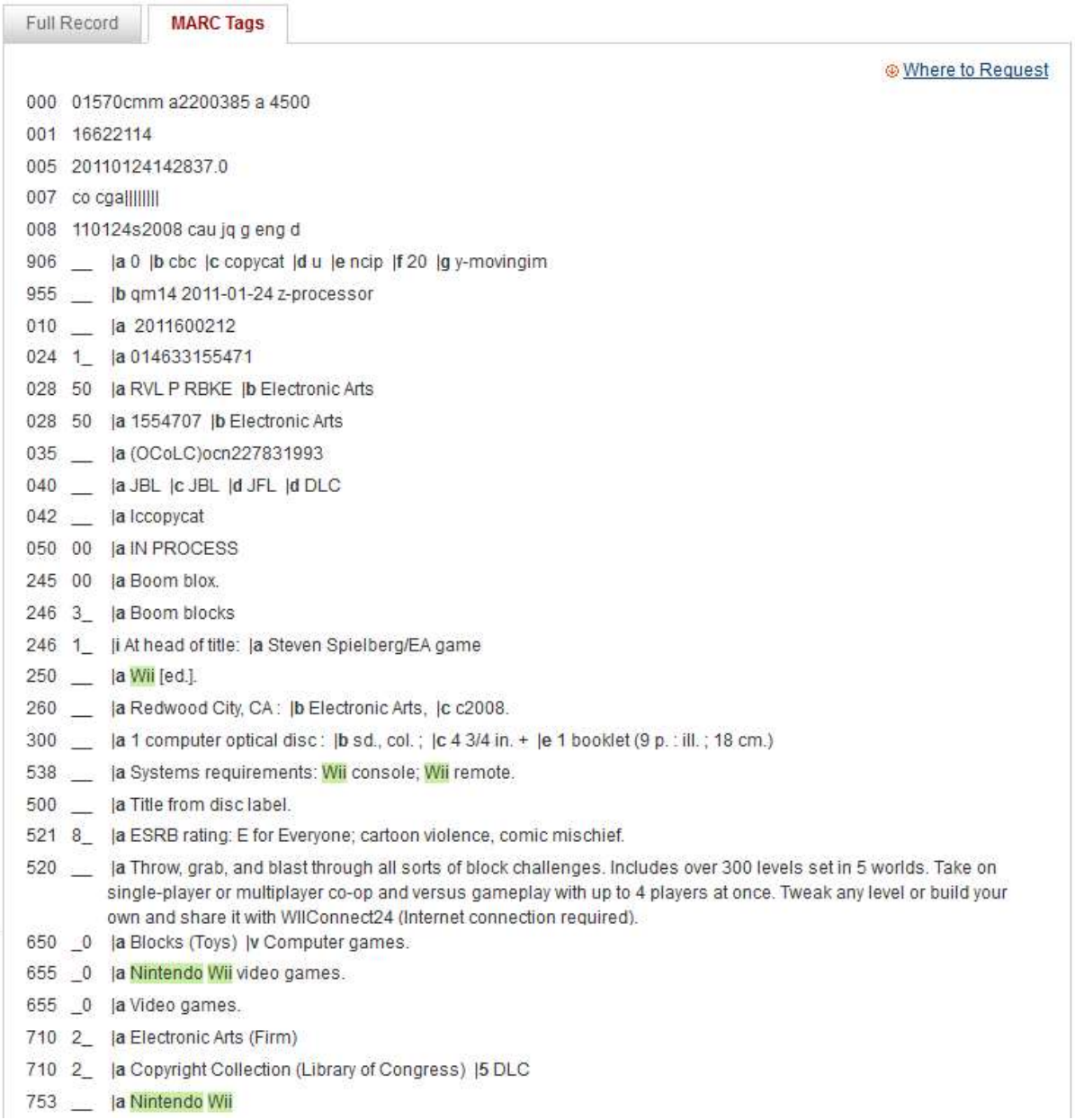
```
Ldr/06: m [type of record: computer file]
Ldr/07: m [bibliographic level: monograph/item]
007:    co#cga|||||||| [category of material: electronic resource
                        specific material designation: optical disc]
008/26: g [type of computer file: game]
245 00: $a World of warcraft $h [electronic resource].
```

En la guía se especifíca que en el campo 008/26 - Type of computer file (006/09)- que "Code "g" indicates that an electronic resource is a game, intended for recreational or educational use. Generally games consist of text and software. A videogame is included here." La LOC hace referencia a los juegos en general incluyendo videojuegos como un recurso recreativo o educacional, con una parte de texto y software entendiendo el mismo como una combinación de audio y video.

En el ejemplo de la LOC (ver Figura 4), se utiliza el campo 028 (“Publisher Number”) para el editor, lo cual es importante porque muchas veces el editor no es el equipo desarrollador siendo preciso hacer mención al mismo. En el caso del ejemplo Electronic Arts es tanto editor como desarrollador, mientras que Ambin Interactive es también desarrollador y no se le hace mención. Los campos 245 y 246 (“Varying Form of Title”) son utilizados para título y a la variación del título respectivamente. En el campo 250 (“Edition Statement”) se hace mención de la edición (y no en el 245 como en el caso de España), que en este caso se corresponde con el sistema. El campo 300 (“Physical Description”) hace una descripción física del soporte; el campo 538 (“System Details”) también da notas sobre detalles del sistema, de la consola y el mando necesarios para el uso de estos recursos. El campo 520 (“Summary, etc.”) presenta una sinopsis sobre el contenido de la obra y el 521 (“Target Audience Note”) la audiencia a la que va dirigido, estos dos aspectos no son contemplados por las bibliotecas españolas.

El registro no incluye número de clasificación (08x). Los campos de puntos de acceso incluyen las siguientes informaciones: el campo 655 además de “video games” también incluye el sistema al que pertenece. el campo 710 (“Established Heading Linking Entry-Corporate Name”) también incluye el nombre del editor/desarrollador Electronic Arts; en el campo 753 también se incluye el nombre del sistema.

Figura 4: registro MARC de recurso electrónico en la Library of Congress



Full Record | MARC Tags

Where to Request

```
000  01570cmm a2200385 a 4500
001  16622114
005  20110124142837.0
007  co cga|||||||
008  110124s2008 cau jq g eng d
906  __  |a 0 |b cbc |c copycat |d u |e ncip |f 20 |g y-movingim
955  __  |b qm14 2011-01-24 z-processor
010  __  |a 2011600212
024  1_  |a 014633155471
028  50  |a RVL P RBKE |b Electronic Arts
028  50  |a 1554707 |b Electronic Arts
035  __  |a (OCoLC)ocn227831993
040  __  |a JBL |c JBL |d JFL |d DLC
042  __  |a lccopycat
050  00  |a IN PROCESS
245  00  |a Boom blox.
246  3_  |a Boom blocks
246  1_  |i At head of title: |a Steven Spielberg/EA game
250  __  |a Wii [ed.].
260  __  |a Redwood City, CA : |b Electronic Arts, |c c2008.
300  __  |a 1 computer optical disc : |b sd., col. ; |c 4 3/4 in. + |e 1 booklet (9 p. : ill. ; 18 cm.)
538  __  |a Systems requirements: Wii console; Wii remote.
500  __  |a Title from disc label.
521  8_  |a ESRB rating: E for Everyone; cartoon violence, comic mischief.
520  __  |a Throw, grab, and blast through all sorts of block challenges. Includes over 300 levels set in 5 worlds. Take on
         single-player or multiplayer co-op and versus gameplay with up to 4 players at once. Tweak any level or build your
         own and share it with WIIConnect24 (Internet connection required).
650  _0  |a Blocks (Toys) |v Computer games.
655  _0  |a Nintendo Wii video games.
655  _0  |a Video games.
710  2_  |a Electronic Arts (Firm)
710  2_  |a Copyright Collection (Library of Congress) |5 DLC
753  __  |a Nintendo Wii
```

La University of Oregon, en Estados Unidos, posee uno de los principales catálogos incipientes de videojuegos (http://janus.uoregon.edu/search/X). En este catálogo también existen registros de hardware (bajo la denominación de material "kit"), el cual se ofrece en préstamo junto al software. Por ejemplo, la consola "Wii" muestra el siguiente registro bibliográfico (ver Figura 5). En el campo 100 ("Main Entry-Personal Name") se incluye el nombre de la compañía, en el 245 ("Title Statement") la denominación y en el 500 ("General Note") todo lo que incluye: la caja completa del hardware, el mando, los cables, cargadores de corriente, manuales, etc.

Figura 5: registro MARC de hardware ("kit") en la Universidad de Oregón

```
LEADER 00000nom  2200000 a 4500
099    Game Kit 3
100    Nintendo
245  0 Nintendo Wii
500    Game kit includes: Wii, 2 remotes, 2 nunchuks, sensor bar,
       AC adapter and cord, AV cable, component video cable,
       eight AA eneloop batteries, battery charger, 3 manuals,
       and case.
```

En el ejemplo de registro bibliográfico de videojuego (ver Figura 6) se observa un uso del campo 260 (“Complex See Reference-Subject”) donde se indica el desarrollador/editor (no el distribuidor), que también se incluye en 037 (“Source of Acquisition”) y en punto de acceso 710 (“Established Heading Linking Entry-Corporate Name”). El resto de campos utilizados son muy similares a los de la LOC, con algunas peculiaridades como el uso del campo 250 (“Edition Statement”) que también denota el sistema, y un uso de notas 500 (“General Note”) en lugar de 520 (“Summary, Etc.”) para número de jugadores y posibilidad de multijugador online.

Figura 6: registro MARC de juego en la University of Oregon

```
LEADER 00000cmm  2200409Ka 4500
001    713955177
007    cb cza|||||||
008    110421s2011    wau        g        eng d
024 10 014633098891
037    BLUS 30732|bValve Corp
037    988907|bValve Corp
040    GL9|cGL9|dGL9|dORU
049    [GAME]ORVN
099    GAME 00262 PS3
245 00 Portal 2|h[electronic resource]
246 3  Portal two
250    Playstation 3
260    Bellevue, Wash. :|bValve Corporation,|cc2011
300    1 computer optical disc :|bsd., col. ;|c4 3/4 in. + 1
       instruction booklet (9 p. : col. ill. ; 15 cm.)
500    Title from disc surface
500    For 1-2 players (network players, 2 players; network
       players co-op, 2). Other features supported: dualshock 3;
       headset
500    Accompanying material may vary
520    Set far in the future from the original Portal game, awake
       in the Aperture Science Labs as Chell, a former test
       subject trying to gain her freedom from the facility and
       GLaDOS, the AI controller of the facility. In escaping
       your initial confinement, GLaDOS is also reawakened. Set
       in the test chambers of the facility, your goal is to make
       your way through the game levels using your portal gun to
       create temporary passages through solid surfaces, allowing
       for creative platforming and multiple possible means of
       clearing a level. Puzzle solving at times is needed to
       clear sections of levels
521 8  ESRB content rating: E 10+, Everyone 10+
538    System requirements: Playstation 3 game system; 24 MB hard
       disk space; supported HD video output: 480p/720P. Optional
       online play requires broadband Internet connection
538    Blu-ray disc; Dolby digital
650  0 Artificial intelligence|vComputer games
655  0 Video games
710 2  Valve (Firm)
753    PlayStation 3
856 4  |uhttp://libweb.uoregon.edu/scilib/games-info.html|zGame
       and Equipment Policies
```

La University of North Carolina también dispone de un catálogo que permite buscar por “video games”, temática y género (http://www.lib.ncsu.edu/catalog/). Para nuestro ejemplo analizamos “Assassins Creed 3” (ver Figura 7), el cual cuenta con el interés añadido de que mediante hechos históricos auténticos se cuenta una historia ficticia, por lo que incluyó un grupo de historiadores para su realización (Rodríguez-

Salces, 2012). Sin embargo, ninguna de estas menciones de responsabilidad son indicadas en la catalogación. En general, los campos utilizados son similares a los de las otras bibliotecas estadounidenses, principalmente a los de la Universidad de Oregón. Como novedad se muestra el uso del campo 082 (“Dewey Decimal Classification Number”) que para el ejemplo es 793.932 (“Computer fantasy games”), lo cual demuestra una denotación del formato y no del aboutness como en la Biblioteca Nacional de España.

Figura 7: registro MARC de juego en la University of North Carolina

Location | Details | Marc Record

```
000 02314cmm a2200553Ia 4500
001 ocn815767846
003 OCoLC
005 20121220053204.0
007 co cga|||||||
008 121102s2012 cau eq g eng d
024 1   |a008888347231
024 1   |a008888397236
028 5 1 |aBLUS-30991 |bUbisoft
028 5 1 |aBPSS-306930 |bUbisoft
028 5 1 |a347231-DISC |bUbisoft
035     |a(OCoLC)815767846 |z(OCoLC)815523881
035     |a(OCoLC)815767846
040     |aJBL |cJBL |dJAI |dWAU
043     |an-us---
049     |aNRCC |xbw
082 0 4 |a793.932 |223
090     |aGV1469.35.A87 |bA873 2012
245 0 0 |aAssassin's creed III |h[electronic resource].
246 3   |aAssassin's creed 3
246 3   |aAssassin's creed three
250     |aPlayStation 3.
260     |aSan Francisco, Calif. : |bUbisoft, |cc2012.
300     |a1 computer optical disc : |bsd., col. ; |c4 3/4 in. + |e1 folded sheet ([10] p. : ill. ; 15 cm.)
500     |aTitle from disc label.
500     |a1 player; 2-8 competitive network players.
500     |aOther features supported: multiplayer, leaderboards, lobbies/matchmaking, messaging/friend invite in
        game, voice chat, add-on content, trophies.
520     |a"The American colonies, 1775. As a native American assassin fights to protect his land and his people,
        he will ignite the flames of a young nation's revolution. Conspire with real historical leaders [and]
        engage in naval warfare"--Container insert.
521 8   |aESRB rating: M, Mature 17+ (blood, intense violence, sexual themes, strong language).
538     |aSystem requirements: PlayStation 3; 4.5 GB hard drive space; 480p/720p/1080i/1080p HD video output;
        broadband required; DualShock 3; headset; 3D game; Blu-ray disc.
596     |a1
650   0 |aAssassins |vComputer games.
650   0 |aQuests (Expeditions) |vComputer games.
650   0 |aRevenge |vComputer games.
650   0 |aHand-to-hand fighting |vComputer games.
651   0 |aUnited States |xHistory |yRevolution, 1775-1783 |vComputer games.
655   7 |aSony Playstation 3 video games. |2NcRS
655   0 |aComputer adventure games.
655   0 |aVideo games.
710 2   |aUbi Soft Entertainment (Firm)
753     |aPlayStation 3
994     |aC0 |bNRC
918     |a2731203
909     |a20121220
019     |a815523881
```

En la Tabla 1 se muestra a modo de síntesis una comparación de las informaciones recomendadas por la ALA que son contempladas por las diferentes bibliotecas analizadas, así como los campos MARC donde se incluyen.

Tabla 1: Información y campos MARC utilizados en los ejemplos de las diferentes bibliotecas

	Comunidad de Madrid	Biblioteca Nacional de España	Library of Congress	University of Oregon	University of North Carolina
Título del videojuego	245	245	245, 246	245, 246	245, 246
Sistema al que pertenece	500	*538	538, 655, 753	250, 538, 753	250, 538, 655, 753
Año de editor/editor	260	260	260	260	260
Clasificación por edades	-	-	521	521	521
Número de jugadores	-	-	520	500	500
Posibilidad de multijugador online	500	-	520	500	500
Sinopsis del juego	-	-	520	520	520
Periféricos específicos necesarios	500	538	538	538	538
Descripción del producto	300	300	300	300	300
Precio	-	-	-	-	-

5 Conclusiones

De los casos analizados se desprende, en sintonía con el informe de la ALA (2011), que no existe una manera normalizada de catalogar este tipo de recursos interactivos. Las bibliotecas estadounidenses generalmente realizan una catalogación más completa que las españolas, ninguna biblioteca incluye todas las informaciones recomendadas por la ALA (como por ejemplo información sobre el precio, que no es incluida por ninguna biblioteca). Las normas de cada biblioteca, la forma en que se cataloga en cada una de ellas y la reciente inclusión de este tipo de recursos interactivos han hecho que la catalogación sea muy dispar, como puede comprobarse con el hecho de que a pesar del uso de catalogación centralizada (copy cataloging) los campos MARC utilizados por cada biblioteca varían. En ocasiones la catalogación depende de la popularidad de un título y de la rapidez con que se haga el registro por parte de la persona que lo cataloga. Si es un juego que no existe en una catalogación previa se debe hacer una catalogación original del mismo (al igual que en la literatura gris existen ciertos juegos sin una popularidad que los condena a la “oscuridad” sin que tengan una catalogación y permanecen como títulos olvidados, es “software gris”). Mientras que muchos vendedores y editores de libros proporcionan una catalogación y clasificación a la biblioteca de los mismos (a través del estándar ONYX), la catalogación previa de los recursos interactivos es en muchos casos nula, por lo que

su aceptación como parte de una colección propia por parte de bibliotecas se ve perjudicada

Referencias

American Library Association. 2011. 2011 **ALA Emerging Leader Team G Final Project Page.** Available http://connect.ala.org/node/146322.

Gee, James Paul. 2007. **Good video games + good learning: collected essays on video games, learning, and literacy**. New York: Peter Lang International Academic Publishers.

Levine, Jenny. 2009. Gaming and Libraries: Learning Lessons from the Intersections. **Library Technology Reports** July: 11-18.

Library of Congress. 2010. http://www.loc.gov/marc/ldr06guide.html#game

McDonough, Jerome, Matthew Kirschenbaum, Doug Reside, Neil Fraistat, and Dennis Jerz. 2010a. Twisty little passages almost all alike: Applying the FRBR Model to a Classic Computer Game. **Digital humanities quarterly** 4 no.2. Available http://www.digitalhumanities.org/dhq/vol/4/2/000089/000089.html

McDonough, Jerome, Robert Olendor, Matthew Kirschenbaum, Kari Kraus, Doug Reside, Andrew Phelps, Christopher Eger, Henry Lowood & Susan Rojo. 2010b. **Preserving virtual worlds: Final Report. National Digital Information Infrastructure Program**. Washington, DC: Library of Congress.

McDonough, Jerome P. 2011. Packaging Videogames for Long-Term Preservation: Integrating FRBR and the OAIS Reference Model. **Journal of the American Society for Information Science and Technology** 62 no.1: 171-184.

Owen, Trevor. 2013. **Video Game Preservation at Scale: An Interview with Henry Lowood.** Available http://blogs.loc.gov/digitalpreservation/2013/02/video-game-preservation-at-scale-an-interview-with-henry-lowood/.

Rodríguez-Salces, Sergio. 2012. **La introducción de los videojuegos en el aula**. Tesis de master. Universidad de La Rioja. http://reunir.unir.net/handle/123456789/766

Sanford, Kathy. 2008. Videogames in the Library? What Is the World Coming To? **School Libraries Worldwide** 14 no.2: 83-88.

SINC/AG. 2009. Desarrollan un emulador para reproducir todos los archivos históricos. Available http://www.agenciasinc.es/Noticias/Desarrollan-un-emulador-para-reproducir-todos-los-archivos-historicos.

Watson, David. 2012. Preserving videogames for posterity. **Multimedia Information & Technology** 38 no.2: 30-31.

Winget, Megan A. 2011. Videogame Preservation and Massively Multiplayer Online Role-Playing Games: A Review of the Literature. **Journal of the American Society for Information Science and Technology** 62 no.10: 1869-1883.

Estrutura de metadados para banco de imagens em patologia

Elisabete Gonçalves Souza
Universidade Federal Fluminense
elisabetes.souza@gmail.com

Jóice Cleide Cardoso Ennes de Souza
UniversidadeFederal Fluminense
E-mail: joicecardoso@id.uff.br

Elan Cardozo Paes de Almeida
Universidade Federal Fluminense
E-mail: elancardozo@id.uff.br

Introdução

O uso de imagens como estratégia de ensino da área biomédica é relevante e contribui para instruir o aluno na definição de diagnósticos, pesquisas e estudos. Especificamente no âmbito dos cursos de graduação da área biomédica, a adoção de recursos visuais como metodologia de ensino é observada há algum tempo, sempre com o objetivo de estimular a construção do conhecimento por parte do discente.

No ensino da Patologia, uma característica fundamental é a necessidade de reconhecer aspectos morfológicos macro e microscópicos (NEVES, 2008). O contato com imagens de lâminas histopatológicas possibilita a reflexão do aluno sobre o conteúdo ministrado em aula, consolidando o conhecimento a partir da explicação de casos clínicos e notícias atuais.

Paralelamente, observa-se a democratização no uso de equipamentos fotográficos, o que estimula a produção de imagens em diferentes contextos. O aumento do número de imagens circulando na sociedade, gerada e arquivada em diferentes equipamentos – *Ipad, smartphones*, etc – leva-nos a questionar sobre como recuperar uma dada imagem a partir dos pontos de acesso existentes. Ampliando esse questionamento para o âmbito das imagens produzidas para uso didático, nos perguntamos quais são os metadados necessários para a descrição de imagens digitais de lâminas histopatológicas com fins de organização e recuperação da informação?

O presente artigo relata a conclusão da primeira etapa do projeto "Organização de imagens de lâminas histopatológicas: abordagem teórico-metodológica visando à recuperação da informação" e objetiva identificar metadados para a descrição de imagens digitais de lesões macroscópicas e de lâminas histopatológicas, produzidas no âmbito didático da disciplina de Patologia Geral, do Departamento de Ciências Básicas (FCB), do Campus Universitário Nova Friburgo, da Universidade Federal Fluminense (UFF); e analisar e contextualizar os metadados para sua inclusão na estrutura *Dublin Core*, para gerar o modelo de metadados para o banco de imagens anatomopatológicas a ser desenvolvido.

Justificativa

A disciplina de Patologia do FCB/UFF disponibilizou nos últimos quatro anos aproximadamente 200 imagens de lâminas histopatológicas para estudo dos alunos dos cursos de Fonoaudiologia, Odontologia e Biomedicina. Atualmente, as imagens estão organizadas de acordo com a disposição dos conteúdos programáticos, não sendo oferecida ao usuário nenhuma forma de recuperação, apresentando somente legenda descritiva. A organização desse acervo, de acordo com os padrões internacionais para a descrição de recursos digitais, permitirá recuperação e acesso das imagens e de todas as informações a elas associadas (tipo de lesão, órgão, diagnostico, etc.), além de favorecer o compartilhamento e a interoperabilidade do futuro banco de imagens com outros sistemas de informação semelhantes.

O estudo é relevante uma vez que as imagens são recursos didáticos importantes nas aulas expositivas e laboratoriais, permitindo que o aluno torne-se sujeito do seu próprio aprendizado pela solução independente (autoinstrução) dos casos clínicos uma vez que acessa as imagens remotamente na página da disciplina.

Tal fato contribui para a ampliação do conhecimento além dos limites de espaço e tempo, enriquecendo o aprendizado e dinamizando-o através do uso de tecnologias atuais, como, por exemplo, a criação de um banco de imagem.

Na revisão de literatura identificou-se a escassez de trabalhos em português que visam o estabelecimento de padrão de metadados para imagens de lâminas histopatológicas. Destacamos o estudo de Carrare *et al* (2006), que desenvolveram ambiente digital de organização e recuperação de imagens em saúde com foco na educação denominado Biblioteca Virtual de Imagens em Medicina (BVIM). Sobre os metadados, os autores esclarecem que "foram utilizados um protocolo predefinido com dados pertencentes às imagens digitais e informações pertinentes aos padrões de registros do sistema de saúde" (CARRARE *et al.*, 2006, p.203), não explicitando o padrão adotado.

Existem **três tipos básicos** de metadados para a organização, a representação e a recuperação da informação: a) os **metadados textuais**, usados na catalogação, cujo objetivo principal é identificar e descrever os recursos informacionais, sendo o padrão mais adequado o AACR2; b) o formato de **metadados simples**, não estruturados, extraídos de forma automática por robôs, como as meta *tags*; e c) os **metadados estruturados** que dão suporte aos motores de busca auxiliando-os na descoberta das informações existentes no ambiente web, como os padrões MARC 21 e *Dublin Core* (ALVES, 2005). Além desses, dispomos de metadados, administrativos, técnicos e de preservação, conforme demonstra a figura abaixo:

TIPO	DEFINIÇÃO	EXEMPLOS
Administrativo	Metadado utilizado na administração de recursos de informação	▪ Aquisição de informação ▪ Direitos de reprodução ▪ Critérios de seleção para digitalização etc.
Descritivo	Metadado para descrição de recursos de informação	▪ Catalogação de registros ▪ Índices especializados etc.
De Preservação	Metadado utilizado para preservação de recursos de informação	▪ Documentação das condições físicas dos recursos etc
Técnico	Metadado utilizado para conhecer as funções de um sistema ou o comportamento dos metadados	▪ Hardware e software ▪ Dados de segurança ▪ Documentação etc.
De uso	Metadado relativo ao nível e tipo de uso de um recurso de informação	▪ Registros de exibição ▪ Sumário de re-uso e de versões etc.

Figura 1 – Tipos de metadados.
Fonte: LOURENÇO (2007).

No âmbito dos estudos das imagens, contamos com o DICOM (*Digital Imaging and Communications in Medicine*) padrão de metadados utilizado mundialmente para a descrição de imagens de diagnóstico (MACHADO, 2002). Conforme Hasegawa e Aires (2007, p.50), a hegemonia do padrão DICOM está relacionada à venda dos equipamentos, sendo ele “um padrão publicado pela Associação Nacional de Fabricantes Elétricos (*National Electrical Manufacturers Association* - NEMA) [que] tem como parceiros mais de 50 indústrias fornecedoras de equipamentos médicos, além de comitês e grupos de trabalho”. Ou seja, a aplicação do padrão DICOM em aplicativos para organização de imagens médicas está diretamente relacionada ao uso desses equipamentos.

O padrão DICOM tem como diferencial o metadado nome do paciente associado à imagem e ao número do registro médico. Contudo, Carro (2003, p.77) esclarece que apesar de disponibilizar esses metadados, o DICOM "não tem conteúdo semântico suficiente para a indexação e pesquisa em contextos de imagens médicas genéricos, como a Web".

Em contrapartida, há o *Dublin Core* (DC), um padrão internacional para objetos digitais em geral, aberto e livre de custos. Além da interoperabilidade semântica, os metadados do DC são passíveis de customização e podem ser adequados às necessidades dos usuários ao descrever seu recurso. No âmbito da descrição de imagens biomédicas seu uso se justifica, pois a utilização de metadados textuais amplia a velocidade da busca pela imagem, tornando-a mais eficiente e precisa. Assim, as similaridades entre as imagens e seus respectivos diagnósticos serão otimizadas, pois ao invés de comparar as imagens, o usuário acessará todos os seus metadados textuais, incluindo o diagnóstico morfológico com qual se relacionam.

Inspirado no modelo DC, Carro (2003) apresenta o modelo de metadados MedISeek para descrição e recuperação de imagens na *web,* baseado em classes, favorecendo a especialização e sua implementação na arquitetura de metadados RDF (*Resource Description Framework*), linguagem de marcação que, utilizando-se da XML (*eXtensible Markup Language*), descreve conteúdos de forma estruturada, permitindo a codificação, o intercâmbio e o processamento de metadados na *web*.

Aires, Matos e Hasegawa (2013) apresentam a utilização do padrão DC para indexação de imagens médicas, com fins de armazenamento, organização e recuperação de dados. Segundo os autores, o padrão proposto também serve como

base na construção do banco de dados, possibilitando a relação entre imagens e metadados previamente definidos.

Metodologia

A partir da análise exploratória e descritiva da literatura existente sobre a aplicação do padrão *Dublin Core* na descrição de imagens biomédicas, elaborou-se um protótipo com os metadados a serem aplicados na descrição de imagens de lâminas histopatológicas.

Conforme Marcondes (2005, p. 96), além de descrever os recursos, o uso de metadados no contexto da *web* permite "agenciar computadores e programas especiais, robôs e agentes de *software*, para que eles compreendam os metadados associados aos documentos e possam recuperá-los". Tal fato oferece ao usuário acesso a uma gama de informações estruturadas, através das quais poderá analisar e avaliar sua relevância, além de manipulá-las, usando-as e reusando-as de acordo com suas necessidades.

A principal característica desta proposta de padrão é permitir a especificidade na descrição de uma imagem através da inserção de novos metadados que contribuam para a descrição integral do recurso de forma contextualizada, tendo em vista seu uso para fins didáticos.

Para a seleção dos metadados a serem aplicados na descrição das imagens de lâminas histopatológicas foi utilizado o padrão DC, por tratar-se de um padrão de descrição para recursos digitais de estrutura simples, objetivo e autoexplicativo permitindo, assim, que o próprio autor de um recurso possa descrevê-lo ao publicá-lo.

Composto por quinze campos, além da simplicidade o DC tem outras características importantes: **flexibilidade:** os elementos que o compõem são opcionais, podendo ser exibidos aleatoriamente e se houver necessidade podem ser repetidos ou modificados por qualificadores; **extensibilidade**: permite que diversas comunidades em áreas distintas utilizem o padrão DC trocando informações e obtendo o acesso a elas; e **interoperabilidade semântica**: diferentes modelos de descrição interferem na habilidade das pesquisas entre áreas, por isso, a aplicação de um modelo comum aumenta a possibilidade de interoperabilidade entre essas áreas.

No processo de criação do padrão de metadados para descrição de lesões macroscópicas e lâminas histopatológicas pesquisamos em repositórios com acervos

imagéticos onde analisamos a estrutura e conteúdo de modo a adequar os padrões observados à proposta do banco de imagens em Patologia. Isso se fez necessário dada a especificidade dessa ferramenta para o estudo e definição de diagnóstico morfológico por imagem, tendo em vista que o conteúdo visual de uma imagem é insuficiente para sua interpretação, sendo necessário descrevê-la integralmente, pois somente com a descrição de informações textuais de forma consistente se tem acesso à "verdadeira expressividade de uma imagem" (CARRO, 2003, p. 15), além de se ter meios para sua recuperação.

Para a definição do modelo foram realizados testes para adequar os metadados DC à complexidade das informações contidas nas imagens biomédicas.

Análise dos resultados

Padrões de metadados como o DC, apresentam uma estrutura de descrição mais formal e detalhada, baseada em normas e códigos especializados de um domínio particular, como os oriundos das ciências documentárias (códigos de catalogação, linguagens controladas, etc.), o que possibilita a descrição da informação e de seus conteúdos de forma mais densa. Isso permite uma recuperação mais precisa e contextualizada dos recursos, como as imagens anatomopatológicas, tendo em vista que o banco prevê campos relacionados a informações diagnósticas, além de outras.

Em uma biblioteca digital, como um banco de imagens, os padrões de metadados incluem os metadados descritivos, que descrevem o recurso informacional (autor, título, assunto, etc.); os estruturais utilizados para estruturação visual da biblioteca digital; e os administrativos, que controlam datas de acesso ao recurso informacional, de direitos autorais etc. (LOURENÇO, 2007).

Nessa direção, propomos o esquema de metadados abaixo para a identificação e descrição das imagens anatomopatológicas produzidas no âmbito da disciplina de Patologia Geral da UFF.

Os elementos foram organizados em: **Propriedade Intelectual** (Figura 2): Criador, Colaborador, Patologista, Editor, Direitos; **Instanciação** (Figura 3): Data de criação, Data de liberação, Data de publicação, Identificador do banco de imagens, Identificador da imagem, Formato da amostra/Ampliação/Coloração, Tamanho/Resolução, Tipo de arquivo, Idioma; e **Conteúdo** (Figura 4): Título, Tipo de

imagem, Descrição, Tipo de lesão, Órgão, Diagnóstico morfológico, Descritores Relação, Forma de aquisição, Responsável pela aquisição, Fonte.

ELEMENTOS *DUBLIN CORE*	NOMENCLATURA	DESCRIÇAO	ORIENTAÇOES DESCRITIVAS
dc.contributor.author	Criador	Nome do criador do recurso.	Entrada de acordo com AACR2, entrada de nomes de pessoas
dc.contributor.advisor	Colaborador	Nome do responsável pela edição da imagem.	Entrada de acordo com AACR2, entrada de nomes de pessoas
dc.contributor.advisor	Patologista	Nome do responsável pelo diagnóstico	Entrada de acordo com AACR2, entrada de nomes de pessoas
dc.publisher	Editor	Nome da instituição responsável pela publicação da imagem.	Entrada de acordo com AACR2, entrada de nomes de entidades. Nesse caso específico, o nome do Universidade/Departamento
dc.rights	Direitos	Informações sobre os direitos de uso.	Uso autorizado para fins didáticos com referência da fonte

Figura 2 - Metadados para a descrição dos **Elementos de Propriedade Intelectual**
Fonte: as autoras

Para a descrição dos metadados referentes às autoridades (pessoas, colaboradores, entidades, etc.) – Figura 2 - o padrão escolhido foi o AACR2 (*Anglo-American Cataloguing Rules*) por ser a norma mais utilizada pelos sistemas de informação no Brasil. O AACR2 tem dois capítulos específicos (Cap. 22 e Cap. 24) para a padronização de cabeçalhos para nomes de pessoas e entidades.

ELEMENTOS DUBLIN CORE	NOMENCLATURA	DESCRIÇAO	ORIENTAÇOES DESCRITIVAS
dc.date.created	Data de criação	Data de criação do recurso	Descrever: dia, mês e ano (dd.mm.aaaa)
dc.date.available	Data de liberação	Data de liberação do recurso para visualização no banco de imagens	Descrever: dia, mês e ano (dd.mm.aaaa)
dc.date.issued	Data de publicação	Data de publicação do recurso no banco de imagens.	Descrever: dia, mês e ano (dd.mm.aaaa)
dc.identifier.banco.uri	Identificador do banco de imagens	URI do banco de imagens	Copiar o link que aponta a URI do banco
dc.identifier.imagem.uri	Identificador da imagem	URI do recurso no banco de imagens	Copiar o link que aponta a URI do recurso
dc.format	Formato da amostra/Ampliação/ Coloração	Descrição física da imagem	Descrever como recurso é apresentado. Ex.: 1 lâmina. Foto em microscopia ótica. Coloração HE – 4x
dc.format.extent	Tamanho/Resolução	Tamanho e resolução do arquivo	Descrever tamanho e resolução em dpi. Ex.: 7 MB – 72 dpi
dc.format.mimetype	Tipo de arquivo	Tipo de mídia e extensão	Descrever o tipo de mídia e a extensão do arquivo. Ex.: image/jpeg
dc.language.iso	Idioma	Idiomas do recurso	Descrever os idiomas nos quais o recurso foi descrito. Ex. por_Br , eng_US

Figura 3 - Metadados para a descrição dos **Elementos de Instanciação**
Fonte: as autoras

Os exercícios de simulação apontaram ser necessário fazer ajustes sendo que alguns campos foram repetidos para atender a integridade da descrição, como os campos *dc.identifier.banco.uri* e *dc.identifier.imagem.uri.* (Figura 3), cujos metadados garantem o acesso ao recurso e ao *site* onde estão depositadas as imagens. Como mencionamos anteriormente, ainda não há banco de imagens, mas sim um projeto para desenvolvê-lo, sendo a arquitetura de seus metadados proposta nesse estudo.

ELEMENTOS DUBLIN CORE	NOMENCLATURA	DESCRIÇAO	ORIENTAÇOES DESCRITIVAS
dc.title	Título	Título do recurso	Descrever conforme consta na legenda da lâmina
dc.type	Tipo de imagem	Natureza ou gênero do recurso.	Descrever mencionando o tipo: Ex.: Fotografia
dc.description.abstract	Descrição	Informações sobre o recurso	Buscar esses dados na legenda do recurso
dc.description.lesaotype	Tipo de lesão	Tipo de lesão identificada	Usar o glossário do DeCs para descrever o termo.
dc.description.orgao	Órgão	Região do corpo em que foi retirado o corte histopatológico	Buscar essa informação na legenda da imagem.
dc.description.diagnostico	Diagnóstico morfológico	Resultado do diagnóstico	Descrever conforme redação feita pelo patologista
dc.subject	Descritor	Assunto principal identificado no recurso	Descrever em português, de acordo com o DeCs
dc.subject	Descritor	Assunto principal identificado no recurso	Descrever em inglês, de acordo com o DeCs
dc.subject	Descritor	Outro assunto identificado.	Descrever em português, de acordo com o DeCs
dc.subject	Descritor	Outro assunto identificado no recurso	Descrever em inglês, de acordo com o DeCs
dc.relation	Relação	Referência a recursos relacionados	Incluir *link* aos quais o assunto se refere. Artigos, etc.
dc.description.origem	Forma de aquisição	Forma de aquisição	Descrever como o recurso foi adquirido. Ex.: Doação
dc.description.origin	Responsável pela aquisição	Pessoa ou instituição responsável pela aquisição do recurso.	Entrada de acordo com AACR2 - nomes de pessoas e/ou entidades
dc.source	Fonte	Tipo de equipamento	Descrever o tipo de equipamento que gerou a imagem

Figura 4 - Metadados para a descrição dos **Elementos de Conteúdo**
Fonte: as autoras

A preocupação com a consistência no processo de busca e recuperação levou-nos a acrescentar os campos *dc.description.lesaotype,* dc.*description.orgaotype e dc.description.diagnostico* (Figura 4). Esses campos são específicos da imagem biomédica e estão relacionados, respectivamente, aos tipos de lesão, à região do

corpo em que foi retirada a imagem macroscópica e corte histopatológico e ao diagnóstico morfológico que auxiliará o clínico na definição do diagnóstico final. Os campos Tipo de lesão, Órgão e Diagnóstico Morfológico são metadados de gestão do banco de imagens e funcionarão como “chaves” para o primeiro corte classificatório, contribuindo para a sistematização de seu conteúdo.

Para a indexação das imagens patológicas - campo *dc.subject* – foi escolhido o vocabulário controlado trilíngue DeCs (Descritores em Ciências da Saúde) por ser referência na área e de fácil acesso pelo usuário. O DeCs é uma iniciativa da BIREME - Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde e foi elaborado a partir do *MeSH – Medical Subject Headings*. Tem como objetivo ser uma linguagem de indexação única de registros de informação da área de saúde e de ser uma linguagem de intermediação na recuperação da informação nas fontes disponíveis na Biblioteca Virtual de Saúde (BIREME, 2015). São 32.160 descritores organizados hierarquicamente, sendo 4.622 exclusivos do DeCs.

Na área da Saúde, o controle dos fenômenos da linguagem natural (sinonímia e homonímia) contribui para o compartilhamento de informações entre instituições. Para tanto, se faz necessária a elaboração de linguagens artificiais levando em consideração a documentação, o usuário e as políticas de informação. As linguagens documentárias são construídas "para indexação, armazenamento e recuperação da informação e correspondem a sistemas de símbolos destinados a ‘traduzir’ o conteúdo dos documentos" (CINTRA *et. al*., 2002, p. 33).

No âmbito da organização e representação da informação os metadados textuais (descritivos e temáticos) garantem a consistência nos processos de busca e recuperação da informação, cabendo aos demais conjuntos de metadados, como os estruturais e os de gestão, manter a interoperabilidade e usabilidade do sistema, permitindo que os usuários possam compartilhar e usar as informações que lhes são disponibilizadas.

Considerações finais

A proposta de criação de um modelo de metadados para o banco de imagens anatomopatológicas confirma a necessidade de se pensar ferramentas para o armazenamento de recursos imagéticos biomédicos que sejam acessíveis aos usuários (professores, patologistas e alunos) de modo que ao depositarem suas imagens possam descrever seu conteúdo de forma objetiva e simples.

A descrição apropriada das imagens biomédicas permitirá ao patologista recuperar diferentes amostras para fazer comparações e aferir seu diagnóstico morfológico. O uso de uma estrutura de banco de dados, como o DC, facilitará a recuperação das informações por descritores autorizados, além de outros dados que garantem a autenticidade de seu conteúdo: nome do criador, curador do banco etc., autoridades cientificas as quais as imagens se relacionam, informações imprescindíveis para o registro dos dados ali depositados.

Para o ambiente da *web* o padrão de metadados DC apresenta-se como uma alternativa confiável e consolidada para a descrição de recursos, como os documentos imagéticos, cujo nível de descrição é mais específico e detalhado.

Como vimos, uma vantagem de sua aplicação é o conjunto de elementos descritivos, que pode ser dividido em dois níveis de estrutura e especificidade: simples, com quinze elementos de descrição, e qualificado, que incluiu elementos adicionais de descrição e indicação de qualificadores, que identificam esquemas para o valor do elemento, como por exemplo, o uso de linguagens documentárias, como o tesauro DeCs, para o elemento "*subject"* e o código de catalogação AACR2, para elementos como "*author"* e "*advisor*".

Todos os elementos de descrição, incluindo os adicionais, elementos de refinamento e esquemas, podem ser repetidos, o que possibilitou idealizarmos para o banco de imagem uma estrutura descritiva rica em informações biomédicas, conforme as necessidades dos estudos de patologia.

Além disso, as imagens terão organização e recuperação adequadas, com a possibilidade de ampliação da recuperação de conteúdo patológico na *web* a partir da propriedade interoperável do padrão *Dublin Core* e seu uso em banco de imagens. A aplicação do padrão de metadados garante uma descrição normalizada e, como consequência, uma representação de qualidade que facilitará o intercâmbio de informações, a interoperabilidade entre os sistemas e, principalmente, a recuperação da informação. Tal fato promove o aumento de estudos de caso no aprendizado da Patologia Geral, dinamizando o processo de ensino-aprendizagem por parte do aluno de graduação.

Pesquisa realizada com auxílio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq

Referências

AIRES, J. P.; MATOS, S.N.; HASEGAWA, F.M. Usando o padrão Dublin Core para definição de metadados de imagens médicas. **Revista Eletrônica FAFIT/FACIC**. Itararé, SP: Faculdades Integradas de Itararé – FAFIT-FACIC, v. 4, n. 2, p. 01-12, jul./dez., 2013.

ALVES, R. C. V. **Web semântica**: uma análise focada no uso de metadados. 2005. 180 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2005.

BIREME. **DeCs** – Descritores em Ciência da Saúde. Sobre o DeCs, 2015. Disponível em: < http://decs.bvs.br/P/decsweb2015.htm>. Acesso em: 17 maio 2015.

CARRARE, A.P. et. al. Uma proposta para gerenciamento e preservação de imagens em medicina na EPM/Unifesp. **Ciência da Informação**, Brasília, v. 35, n. 3, p. 201-208, set./dez. 2006.

CARRO, S.A. **Um modelo de metadados para a indexação e recuperação de imagens médicas na web**. Dissertação (Mestrado em Ciência da Computação). Porto Alegre, 2003.

CINTRA, A. M. M. et.al. **Para entender as linguagens documentárias**. 2. ed. São Paulo: Polis, 2002.

HASEGAWA, F.M.; AIRES, J. P. Proposta de um Padrão de Metadados Para Imagens Médicas. **ERI** - Escola Regional de Informática, Guarapuava, PR, 2007.

LOURENÇO, C.A. Metadados: o grande desafio na organização da web. **Informação & Sociedade**, João Pessoa, v.17, n.1, p.65-72, jan./abr., 2007.

MACHADO, M. S. **Modelo de metadados para armazenamento de recuperação de imagens estáticas no formato DICOM**. PPGC da UFRGS, 2002. Dissertação de mestrado Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação.

MARCONDES, C.H. Metadados: descrição e recuperação na Web In: ____ et al. (Org.). **Bibliotecas digitais:** saberes e práticas. Salvador, BA: EDUFBA; Brasília; IBICT, 2005. p. 77-143.

NEVES, F.B.C.S. et al. Impacto da Introdução de Mídia Eletrônica num Curso de Patologia Geral. **Revista Brasileira de Educação Médica**, v. 32, n. 4, p. 431-436, out./dec.2008.

UNIVERSIDADE FEDERAL FLUMINENSE. Faculdade de Odontologia de Nova Friburgo. **O que é patologia.** 2015. Disponível em: <http://patogeralpunf.wix.com/punfuff#!sobre/c1t44>. Acesso em: 17 maio 2015.

Indexação por imagens: acessibilidade via OPACs imagéticos

Rita de Cássia do Vale Caribé
Universidade de Brasília
rita.caribe@gmail.com

Marcílio de Brito
Universidade de Brasília
mdebrito@unb.br

Introdução

A sociedade da pós-modernidade é caracterizada por um esforço de modernização compulsiva e obsessiva, pela complexidade e incerteza. Nesse contexto a informação tem papel relevante, pois permite que os indivíduos possam se apoiar na informação de qualidade para satisfazer suas necessidades de informação, seja para o exercício profissional, para o exercício da cidadania ou para suas atividades cotidianas.

O processo de geração do conhecimento está disperso entre uma grande variedade de instituições e pessoas que sistematizam informações com o objetivo de facilitar a sua recuperação e localização do documento primário quando necessário. Em lugar privilegiado estão as bibliotecas, unidades sociais caracterizadas pela organização de conjuntos de objetos informacionais objetivando sua recuperação e disseminação em sintonia e sincronia com as demandas efetivas e potenciais de sua clientela.

No processo de organização da informação, especificamente a etapa de descrição do conteúdo do documento, a linguagem é utilizada para representar os conceitos que nele são tratados.

A questão da representação do conteúdo se torna complexa quando se lida com linguagens pouco comuns, especialmente aquelas usadas por indivíduos com necessidades especiais ou deficientes.

Entende-se igualmente que a natureza dos acervos é tão diversificada quanto a natureza dos documentos, assim ao resolver-se o problema do acesso ao

documento, o processo de comunicação documento-usuário se estabelece pela própria natureza de suas afinidades. O marco importante é colocar a informação em contato com seu usuário. Partindo-se do pressuposto que o catálogo das bibliotecas é a porta principal pela qual se busca localizar os documentos em um acervo qualquer, pergunta-se como dar acesso aos catálogos de bibliotecas?

A descrição de conteúdo no sentido tradicional

Tradicionalmente, no processo de análise de conteúdo identificam-se os assuntos e a forma sob a qual foram tratados, suas contiguidade temporal e espacial dentre outros aspectos relevantes para facilitar a localização do documento quando procurado.

A operação denominada indexação, é o “ato de identificar e descrever o conteúdo de um documento com termos representativos dos seus assuntos e que constituem uma linguagem de indexação” (ABNT, 1992, p. 2). O UNISIST (1981) complementa especificando que a indexação torna possível recuperar os conceitos contidos nos documentos que podem ser representados por termos selecionados da linguagem natural (palavras-chave), de um tesauro (descritores) ou por símbolos (número de classificação retirado de um sistema de classificação bibliográfica como a CDD ou CDU).

No processo de indexação os conceitos são extraídos do documento por meio da análise, portanto, em linguagem natural, rica em significados e figuras, como metáforas, licenças poéticas, entre outros. A linguagem natural é caracterizada pela conotação.

A operação que permite utilizar as palavras-chave da indexação, consiste em transformar conceitos em termos, os mesmos encontrados nos tesauros, listas de cabeçalhos de assunto, sistemas de classificação, etc.

As abordagens linguísticas deram seus frutos (BRITO, 1992) e comprovaram, mais recentemente com o progresso da tecnologia, a propriedade deste procedimento. Apesar disso, métodos simplificados de metadados, como os *tags* das folksonomias, se propagam consideravelmente nos sistemas de armazenamento e recuperação da informação. A razão para isto é a simplicidade da implementação, embora conhecidos os efeitos de revocação e pertinência na recuperação.

Nem tudo, portanto é ruim e nem tudo é eficiente por completo. Mesmo os modelos linguísticos mais sofisticados, dos analisadores (*parsers*) do texto integral, confrontam-se a desvios de performance devido à grande flexibilidade das leis que regem o discurso produzido pelas comunidades, em contraste com a rigidez das gramáticas de suas línguas.

Nesse processo de evolução das ferramentas de tratamento da informação escrita, verificou-se relativa evolução no reconhecimento da forma (morfologia) e das relações sintáticas inter e entre unidades do discurso. Os sistemas baseados principalmente em estatísticas tomaram proveito deste avanço atribuído aos métodos linguísticos e viu-se o aparecimento dos sistemas ontológicos e da web semântica.

O uso todavia desses métodos simplificados, híbridos e estatísticos, em oposição ao proposto pelos métodos linguísticos, devolve ao usuário a competência para identificar a pertinência dos resultados. Aceita-se passivamente como normal a imprecisão dos sistemas e a incerteza na exaustividade da busca.

A problemática posta é aquela em que o momento tecnológico pede novas soluções sem alterar o balanço desta relação custo-benefício. Até que ponto é possível modificar a cadeia de produção (indexação da informação), sem abalar as atuais instituições documentárias, e simultaneamente aumentar a satisfação do usuário ?

Estudos de marketing (ASHTON, 2015) demonstram as vantagens do uso de imagens e levantam considerações importantes sobre a mídia informacional. Saber ler é essencial, mas nem todo mundo sabe ler da mesma forma. A informação registrada não é apenas a informação escrita, a imagem conquistou um espaço nesse processo de comunicação.

A semiologia mostra que a imagem é uma linguagem de comunicação completa, e nesse particular as competências “linguísticas” da população, amplamente consumidora de imagens, é muitas vezes superior àquelas da língua escrita. As leis e códigos de comunicação que regem a linguagem imagética são, por conseguinte, mais compreensíveis para o grande público do que os tradicionais códigos escritos.

A leitura de imagens supera a performance leitura da mensagem escrita com maior precisão e por um público mais amplo:

- Os cérebros consomem 1/10 segundo para perceber uma imagem de paisagem em oposição aos 60 segundos necessários à leitura da descrição dessa mesma cena (200-250 palavras).
- Imagine-se viajando a 80 Km/h e lendo o texto a seguir, “Adverte-se ao condutor do veículo da existência adiante, de área sujeita a desmoronamento ocasionado por instabilidade no talude ou por obras”, em vez desta imagem .
- A inclusão de vídeo em mensagens no Twitter aumenta os Retweets de 28%, e tweets que incluem fotos atingem 35% mais Retweets (ROGERS, 2014);
- Um estudo publicado por Social Bakers (2014) mostrou que fotos representam 87% dos posts compartilhados em 30.000 páginas, enquanto qualquer outro tipo de conteúdo não atinge mais do que 4% dos compartilhamentos.

Introduzindo um novo conceito de OPAC

Diante das evidências desta preferência (humana) pelo uso de imagens nos sistemas de comunicação, pergunta-se, por que as bibliotecas não adotariam essa forma de comunicação? Afinal de contas, comunicar-se não faz parte de seus valores culturais vinculados às ações cidadãs e para a inclusão social, favorecendo o acesso à informação e à documentação como parte da universalidade dos direitos humanos ? Esta não seria uma via de evolução natural dos sistemas atuais ?

Essas questões evoluem em vários sentidos, por um lado considerando as características dos usuários, por outro, estudando as adaptações técnicas e funcionais dos processamentos bibliográficos, dos produtos e serviços das bibliotecas.

Com a utilização de tecnologias da informação, os chamados *Online Public Access Catalogue* (OPAC), dominam atualmente o cenário das bibliotecas. A população letrada acessa os OPACs por meio de interfaces essencialmente baseadas na linguagem escrita. O usuário se comunica com o acervo da biblioteca por meio de menus, listas alfabéticas ou campos de busca do estilo Google. Embora menos frequente, o público infantil dispõe de interfaces gráficas adaptadas às pequenas coleções. À margem desse processo encontram-se as comunidades minoritárias tais como os iletrados, tanto digitais como os funcionais, os deficientes visuais e auditivos,

os autistas, os idosos dentre outros. Se por um lado os deficientes visuais possuem programas que fazem a audio-leitura das telas de computador, poucas ferramentas estão disponíveis para compensar essas barreiras dos demais.

A reformulação do conceito atual de catálogo, baseado na linguagem escrita, para um novo conceito de catálogo imagético suscita respostas pela emergência das novas tecnologias, mas confronta-se com outros questionamentos referentes à epistemologia dos catálogos, às características e propriedades das imagens.

Referenciais linguísticos para o processamento da informação

O processo de busca de informações leva à constatação de uma relação certamente complexa entre escrever, comunicar e descrever a língua, para expressar o pensamento. De fato, o que se deseja é poder representar os conhecimentos materializados, por atos de linguagem, em um texto. Esta representação passa pela aplicação de métodos capazes de veicular propriedades intrínsecas à língua e à sua gramática. A pesquisa que se empreende aqui consiste, substancialmente, em transpor os conteúdos documentários em uma meta-linguagem utilizando-se de imagens como representantes dessa (meta) linguagem controlada.

Na França dos anos 1960, os linguístas se aplicaram nas pesquisas descritivas. Os problemas até então limitados, foram tratados sob o ponto de vista dos “campos morfo-semânticos” ou sob a perpectiva lexicológica dissociada da morfologia, trabalhando-se sobre corpus circunscritos: descrição semântica do vocabulário da habitação, animais domésticos da moda entre outros. Pouco a pouco adere-se às propriedades físicas dos objetos denotados pelas palavras estudadas, certo progresso se dá a partir do exame dos códigos ou linguagens documentárias destinadas à recuperação dos objetos sobre os textos de origens arqueológicas. Na pesquisa em teoria semântica destacam-se, principalmente, as pesquisas da linguística sobre a respresentação da informação (BARTHES, 1967; GREIMAS, 1966, 1970, 1973, 1976; GREIMAS, s.d.; GUIRAUD, 1966; MOUNIN, 1965) que deram origem às pesquisas sobre a representação da linguagem natural e em particular ao seu processamento automático, levando-se em conta os numerosos componentes que intervêm nesse processo – morfológico, lexical, sintático, semântico, e lógico – assim como as relações e dependências entre os elementos simples e complexos do discurso.

Sabe-se, todavia, que em algum lugar do texto existe uma significação (objetos que inferem em conhecimentos), que se pode localizar por operações naturais, tais como a leitura, mas que não se pode representar convenientemente de forma artificial.

O critério essencial de modernidade nos métodos de tratamento da informação não reside no emprego de equipamentos sofisticados e modernos, mas na adoção de uma nova forma de se apresentar o problema. A análise de um documento prevê que este seja tratado convenientemente por um sistema documentário. Essa operação consiste em indexar o documento, descrever seu conteúdo informacional, respeitando-se as limitações impostas pelas linguagens documentárias. A indexação, tal como é conhecida, é uma tradução lexical de unidades da língua, ou ainda uma tradução sintática quando se trata de expressar as ligações sintáticas entre as diferentes partes do discurso, aquelas que descrevem o conteúdo, os descritores.

Segundo o uso que se pode fazer das palavras, elas podem ser de quatro tipos distintos:

- Palavras da língua
- Palavras do léxico
- Palavras do discurso, e
- Palavras em terminologia (Termos)

A indexação é uma operação que visa a representação dos objetos de que fala o documento. Em linguística, esses objetos são chamados de referentes. Assim, as palavras da língua ou do dicionário (morfemas) remetem unicamente aos seus significados, e não se pode dizer portanto que elas apontam para referentes.

As palavras do léxico, lista de palavras de um sistema documentário, não possuem o status de palavras da língua, nem tampouco de palavras do discurso. As palavras do léxico não designam objetos, elas remetem a um conjunto aberto de objetos que possuem características comuns. É impossível delimitar precisamente as fronteiras deste conjunto, e por conseguinte designar um referente, para se estabelecer uma conexão direta entre as palavras do léxico e um objeto da realidade (realidade extra-linguística ou realidade imaginária). A sutileza que separa as palavras do léxico das palavras em terminologia explicam e justificam a tendência a confundir esses dois termos.

As palavras em terminologia introduzem a questão de fronteira. Como explica o linguista Michel Le Guern (1989), no léxico como em terminologia encontram-se "palavras", mas de um lado como de outro, não são as mesmas palavras. O objeto "palavra" do léxico é uma realidade distinta, o léxico trata as palavras independentemente das coisas, enquanto em terminologia, as palavras estão ligadas às coisas. As palavras em lexicografia designadas como substantivos são em realidade predicados. Elas falam das qualidades e não das substâncias. As palavras do léxico remetem a propriedades e não a substâncias, a qualidades e não a objetos. A existência de um objeto demanda um termo (palavra em terminologia) para entrar no discurso.

É somente pela aplicação no discurso e pela sintaxe que as palavras do léxico podem participar na construção dos termos, i.e. entrar em relação direta com as coisas (objetos, referentes).

Esta é uma operação da lógica, e em lógica as palavras do léxico não possuem *extensão*, elas possuem apenas *intensão* (compreensão). Ao discurso corresponde uma lógica *extensional* e às palavras da língua uma lógica *intensional*, e uma se opõe à outra. A lógica extensional constrói segmentos que são portadores de referência a uma realidade, que designa objetos do universo. Fala-se, portanto de classes de objetos, deve-se considerar um universo constituído de objetos. A lógica intensional, das palavras da língua, aplica-se às noções (predicados) que fazem referência aos conceitos.

O fato que se tenha no léxico predicados livres e em terminologia predicados vinculados implica que pela lógica combinatória é possível se ter predicados complexos de ambos os tipos. Também pelas operações da lógica, basta que um dos predicados seja vinculado para que o predicado complexo também o seja.

A situação que consiste em descrever o conteúdo dos documentos com palavras-chave do léxico parece assim pouco sustentável, mesmo admitindo-se a existência de um código (a linguagem documentária) pretensamente universal, para assegurar a transição léxico-discurso. É preciso buscar no discurso os elementos necessários para a representação dos temas do discurso.

De posse dessas informações, em que se esclarece a natureza do signo, representante das referências da realidade extra-línguística, i.e. dos objetos do discurso, entende-se que a qualidade dos sistemas de recuperação está intimamente

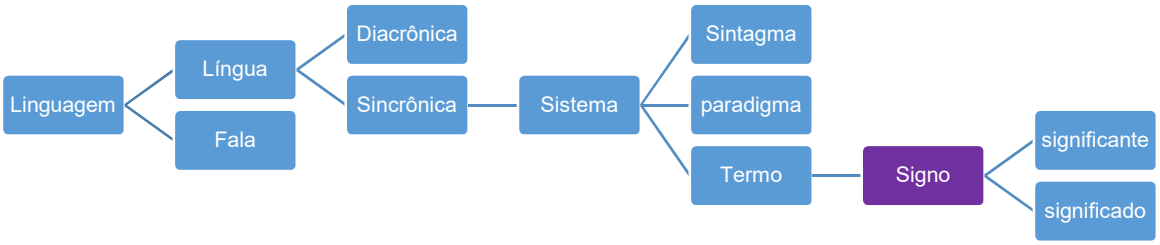
ligada às propriedades de vinculação entre unidades do discurso e objetos. Esta é a essência dos mecanismos de busca documentária. Em resumo, a eficiência dos sistemas de recuperação de informação está intimamente relacionada à identificação das coisas de que tratam os documentos. As linguagens documentárias devem, pois assumir esta função de apontar com pertinência para os objetos, e situar o leitor neste universo contextual.

Retoma-se, dessa forma, as questões postas anteriormente, sobre o uso de imagens nos catálogos, que por sua vez demandam uma nova linguagem documentária de mesma natureza, i.e. baseada em imagens.

A estrutura teórica de suporte

As duas ciências Semiologia, desenvolvida na Europa por Ferdinand Saussure linguista frances, e Semiótica, desenvolvida nos Estados Unidos pelo filósofo americano Charles Sanders Peirce, ambas estudam os signos na vida social, embora sejam escolas diferentes. A semiótica provê conceitos e idéias que permitem uma análise sistemática dos signos, que podem ser palavras, imagens, textos completos, artefatos, gestos, posturas físicas etc., estuda ainda o meio em que os signos são utilizados e como os significados são construídos.

Figura 4 – O signo e a linguagem segundo Saussure: representação esquemática.



Paralelamente e na mesma época (1900), aparecem também em semiótica, as teorias da indexicalidade de conteúdos de/por imagens. A revisão dessa abordagem converge para a teoria do signo de Charles Sandes Peirce, que defende uma semiótica fundamentada na relação entre o símbolo, o ícone e o índice. Linguagem e imagem compõem-se assim de signos que possuem uma função icônica, indexal e simbólica. Para maior objetividade do assunto aqui tratado, o nível de simplificação adotado para a teoria de Peirce estará centrado na tríade índice, ícone e símbolo, da seguinte forma:

- O ícone é um signo em que o significante corresponde ao significado, por exemplo, a imagem de uma caneta em uma caixa indica que dentro encontram-se canetas do mesmo tipo. A imagem é uma representação fiel da realidade.
- O índice é um signo representado pela resultante que o significado causa no significante. Assim, pessoas com guarda chuva na rua é um indício de que está chovendo. A representação da chuva que não se vê na imagem passa pela expressão do(s) ícone(s).
- Um signo simbólico existe quando a relação entre significante e significado é puramente convencional, tal como as placas de trânsito, tal como a pomba que representa a paz.

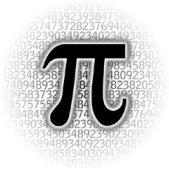
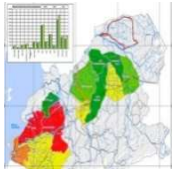
Cabe ressaltar que a teoria semiótica de Peirce é amplamente aplicada à descrição de imagens (BOULANGER, 1996). Nas coleções imagéticas e nas práticas arquivísticas, trata-se de um processo de descrição literal de um documento imagético para efeito de recuperação.

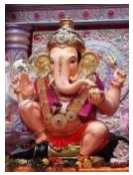
Vê-se, assim, que a representação de conteúdos através de imagens pressupõe uma operação diferenciada da indexação. As vantagens que se anunciam são múltiplas: atração visual do usuário pelo sistema, compreensão intuitiva do código de indexação, ampliação da portabilidade dos indexadores, interoperabilidade entre as leis que regem os universos do discurso e do profissional indexador e em consequência, também com o universo de compreensão do usuário.

Propõe-se que as bases metodológicas para este conceito de “catálogo online imagético de biblioteca” sigam as práticas atuais e características das bibliotecas. Assume-se a adoção dos sistemas de classificação internacionais: a Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU).

Para efeito do presente exercício adotou-se a CDU como exemplo. A CDU é uma classificação baseada em 10 classes que se subdividem cada uma em 10 sub-classes e assim sucessivamente de acordo com a especificidade de cada área. O quadro a seguir apresenta, a título experimental e ilustrativo, a representação das classes da CDU.

Quadro 2 - Representação gráfica das 10 classes da CDU

0	1	2	3	5	6	7	8	9
CIÊNCIA E CONHECIMENTO. ORGANIZAÇÃO. INFORMÁTICA. INFORMAÇÃO. DOCUMENTAÇÃO. BIBLIOTECONOMIA. INSTITUIÇÕES. PUBLICAÇÕES	FILOSOFIA. PSICOLOGIA	RELIGIÃO. TEOLOGIA	CIÊNCIAS SOCIAIS	MATEMÁTICA. CIÊNCIAS NATURAIS	CIÊNCIAS APLICADAS. MEDICINA. TECNOLOGIA.	ARTE. RECREAÇÃO. ENTRETENIMENTO. DESPORTO	LÍNGUA. LINGUÍSTICA. LITERATURA.	GEOGRAFIA. BIOGRAFIA. HISTÓRIA
 01 - Bibliografias e catálogos	 11 - Metafísica	 21 - Religiões pré-históricas e primitivas	 31 - Estatística	 51 - Matemática	 61 – Ciências médicas	 71 - Planejamento territorial, físico. ... regional, urbano e rural...	 81 - Linguística e línguas	 91 - Geografia. Exploração da terra e de países. Viagens. Geografia regional
 02 – Biblioteconomia		 22 - Religiões do Extremo Oriente	 32 - Política	 52 - Astronomia	 62 – Engenharia. Tecnologia em geral	 72 - Arquitectura	 82 - Literatura	 92 - Biográfias. Genealogia. Heráldica. Bandeiras

								
03 - Obras gerais de referência	13 - Astrologia	23 – Religiões do sub-continente indiano. Hinduísmo	33 - Economia	53 - Física	63 – Agricultura, caça e pesca	73 - Artes plásticas		93 - História
	14 - Psicologia	24 - Budismo	34 - Direito. Jurisprudência	54 - Química	64 - Economia doméstica. Ciência doméstica	74 - Desenho. Design. Artes e ofícios aplicados		
05 - Publicações periódicas, periódicos		25 - Religiões da antiguidade. Cultos e religiões menores	35 -Administração pública. Governo. Assuntos militares	55 - Ciências da terra. Ciências geológicas	65 - Indústrias da comunicação e dos transportes. Contabilidade. Gestão de empresas. Relações públicas	75 -Pintura		
	15 – Filosofia							
06 Organizações em geral 070 Jornais. Imprensa 08 Poligrafias. Obras de autoria colectiva 09 Manuscritos. Obras raras e notáveis	16 Lógica. Epistemologia. Teoria do conhecimento. Metodologia da lógica 17 Filosofia moral. Ética. Filosofia prática	26 Judaísmo 27 Cristianismo. Igrejas e denominações cristãs 28 Islamismo 29 Movimentos espirituais modernos	36 - Protecção das necessidades materiais e mentais da vida. Serviço social. Ajuda social. Segurança social. Habitação. Consumo. Seguros 37 – Educação	56 Paleontologia 57 Ciências biológicas em geral 58 Botânica 59 Zoologia	66 Tecnologia química. Indústrias químicas e relacionadas 67 Indústria, artes industriais e ofícios diversos 68 Indústrias, artes e ofícios de artigos acabados ou montados 69 Indústria da construção.			

			39 - Antropologia cultural. Etnologia. Etnografia. Usos e costumes. Tradições. Modo de vida. Folclore		Materiais para construção. Procedimentos e práticas de construção			

Nota: Imagens de domínio público. Tabela parcialmente ilustrada para efeito de estudo.
Fonte: Adaptado de Biblioteca pública. Governo dos Açores

Nessa interface gráfica, a busca do uuário se daria pela escolha (facetada) hierárquica de imagens, ela propõe uma navegação em cadeia pelas sub-classes do catálogo (CDU) até a obtenção dos documentos desejados.

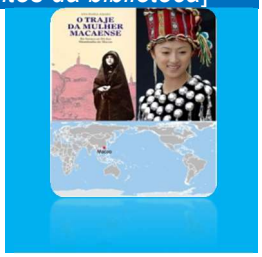
Então, se o usuário procura um documento sobre “Vestuário. Indumentária. Traje nacional. Moda. Adornos (sub-classe CDU 391), o procedimento seria o seguinte:

- O usuário percorre a página do catálogo em imagens até encontrar seu assunto representado na coluna da classe 3 (Ciências Sociais), sub-classe 39.

Quadro 5 - Representação gráfica da sub-classe 39 da CDU

Fonte: elaboração própria (imagens de domínio público)

Quadro 6 - Representação gráfica da sub-classe 391 da CDU

Guia de navegação	Classe 391 - Vestuário. Indumentária. Traje nacional. Moda. Adornos *[grupo de imagens representativas da classe e dos documentos da biblioteca]*
 ⇓ sub-classe 39 ⇓ sub-classe 391	

Fonte: elaboração própria (imagens de domínio público)

Figura 7 - O resultado da busca contém a referência da obra no sistema e sua imagem (carta-gráfica)

Reference
» Traje da mulher macaense : Da saraca ao do das nhonhonha de macau(o) / 1989
TRAJE da mulher macaense: Da saraca ao do das nhonhonha de macau(o).
O TRAJE DA MULHER MACAENSE
Macao
See also
Content data | Copies | Location | Reference | MARC | Reservation

Fonte: elaboração própria (imagens de domínio público), a partir da tela de resultado do software Pergamum da BU.

Nesse procedimento, em que o usuário percorre a lógica hierárquica do catálogo imagético ou realiza uma busca aleatória por imagens, o resultado final deve ser a referência do documento, ou conjunto de documentos semelhantes em imagens e em texto, os quais darão acesso à localização física dos documentos no acervo.

A esta representação simplista de navegação no catálogo imagético pressupõe uma argumentação teórica subjacente. Segundo a teoria pragmática de Peirce, cada objeto do universo se relaciona com um indeterminado número de outros objetos, tanto direta como indiretamente. De tal forma que cada elemento transporta um potencial indicial indeterminado (LEFEBVRE, 2007, p. 220). Como então o signo representa seu objeto, e como ele revela seu objeto? A representação icônica aparece como uma qualidade que possui o objeto, se essas qualidades emergem, então o objeto pode ser identificado como tal e através do seu ícone. O caráter indicial está intimamente ligado ao objeto: o frio ou temperaturas baixas estão intimamente ligados à imagens de neve ou de gelo. Simbolicamente, o signo precisa ser interpretado. É pela ação dos signos que se infere que uma realidade é verdadeira e neste contexto pode-se participar da indexação de imagens. Pelo conhecimento, o signo pode ser interpretado, ampliando-se os limites da relação signo-objeto.

Considerando-se a tarefa de leitura da imagem, as teorias saussuriana e peirciana seriam suficientes para que, com a prática adequada, um profissional da indexação pudesse indexar imagens atribuindo-lhe palavras-chave segundo seus níveis de percepção: significante e significado, conotação e denotação, ícone, índice ou símbolo. O vetor resultante desses indicadores serviriam para armazenar e recuperar o documento imagem adequadamente. Mas, a inovação que se deseja agregar ao processo é justamente substituir o papel do signo palavra-chave tradicional, por um signo imagem, que representaria o conteúdo dos documentos. Sobre isso, a contribuição de Jacques Bertin (1967 *apud* DANTIER, 2008; BERTIN, 1970) é particularmente enriquecedora.

Sobre a análise da informação, Bertin esclarece que todo pensamento se expressa por um sistema de signos, imitando uma codificação natural. A linguagem verbal é um código de signos sonoros, a escrita de uma língua é outro tipo de código, e a representação gráfica também. Dado que a representação gráfica é a transcrição da informação em um sistema gráfico de signos, a representação gráfica faz portanto parte da semiologia.

Transcrever significa separar os elementos do pensamento qualquer que seja o sistema de signos disponível. Independente da qualidade dessa transcrição, somente interessa a qualidade e a eficácia da transcrição gráfica, pois é esta que representa o conteúdo da informação. Cada sistema de signos dispõe de seus próprios meios, estilo e estética para representar o pensamento. Trata-se de uma relação entre diversos conceitos reconhecidos e isolados, escolhidos dentre todos os conceitos imagináveis. Assim, diz Bertin (*apud* DANTIER, 2008, p. 19, tradução nossa, grifo do autor) que

> [...] na representação gráfica, chamar-se-á INFORMAÇÃO o conteúdo traduzível de um pensamento. Ele [o conteúdo] constitui-se essencialmente de uma ou várias CORRESPONDÊNCIAS ORIGINAIS dentre um conjunto finito de conceitos de variação, e um conjunto invariante.

A teoria da imagem de Bertin introduz também o conceito de eficácia da imagem que se define em função do tempo de percepção da mensagem. Para se obter uma resposta correta e completa a uma dada questão, a "construção" que requer o tempo mais curto de percepção da resposta é a construção mais eficaz, e corresponde à imagem. Esta noção de "custo mental" foi posta em evidência pela lei de Zipf (1949), muito usada em Biblioteconomia.

Em sua teoria da imagem, Bertin evidencia três operações sucessivas para a leitura de uma imagem que vão guiar posteriormente o processo inverso de construção de imagens:

1. Identificação do externo – concretizar no pensamento os conceitos propostos, os componentes.
2. Identificação interna – por meio de que variáveis os componentes se expressam.
3. Percepção das correspondências originais. Esta percepção é o resultado de uma questão, consciente ou não. A questão que se pode colocar diante de uma informação.

Isto significa que o olho pode isolar as informações visuais pelo reconhecimento das correspondências definidas e perceber a mensagem em seu contexto. Esse exercício de percepção conduz à definição de imagem como sendo "a forma visual significativa perceptível no instante mínimo de visão".

Finalmente, retém-se como justificativas para a composição de uma representação gráfica, as três funções básicas: registrar, comunicar e tratar a informação.

O processo de indexação **por** imagens prevê assim um momento de leitura e outro de concepção. Indexar por imagens consiste em construir uma imagem que represente o conteúdo do documento.

- Ferramentas equivalentes às linguagens documentárias:
    - Deve-se constituir um banco de imagens preferencialmente padronizadas segundo uma política de programação visual estabelecida.
    - As mesmas remissivas dos instrumentos tradicionais (Imagem genérica, imagem específica, Use, Usado para, Ver etc.) devem ser previstas para as imagens, estabelecendo-se padrões para o uso das imagens (GHEORGHITA, 2011).
    - Os princípios anunciados por Bertin (*apud* DANTIER, 2008) devem ser observados para dar ao corpus de imagens uma identidade e um estilo de linguagem.
    - Um banco de imagens pode, tal como um tesauro, ser construído por profissionais capacitados, a fim de se obter uma coleção homogênea e coerente com a área de especialização.
    - Às imagens desse banco de imagens serão associados metadados para facilitar a recuperação das imagens e a transmissão de suas características temáticas às colagens (imagens compostas);
- O resultado da indexação:
    - Contrariamente à palavras-chaves, o produto não será uma seleção de imagens justapostas, mas uma colagem de imagens significativas, de forma a gerar uma “carta gráfica” da obra. Esta carta gráfica será o signo simbólico da obra, herdando todas as características das imagens que a compõem.

- A composição desta “carta gráfica” se dará pela composição mínima de um signo icônico e um signo indicial, fechando-se a tríade peirciana, visto que a própria “carta” pretende ser um signo simbólico.
- A escolha do signo icônico deve ser, tanto quanto possível, a imagem da capa da obra, ou aquela que melhor a representa (há livros que não possuem imagens na capa).
- A composição da carta gráfica da obra continua com a colagem dos signos indiciais e simbólicos, acessórios que permitem detalhar o conteúdo da obra.
- Como, pelas dimensões reduzidas, a carta gráfica é composta de poucas imagens, deve-se dar prioridade a uma imagem de localização geográfica e outra temporal, situando a obra nesses dois eixos lógicos.

O experimento de indexação por imagens apresentado a seguir, é apenas ilustrativo do que se pode obter como resultados. Pelo fato de não se ter constituído o banco de imagens padronizadas, equivalentes à linguagem documentária tradicional, usou-se imagens de domínio público recuperadas na web.

Escolheu-se aleatoriamente três obras reais recuperadas no catálogo da Biblioteca Universitária (BU), dentro da classe CDU (39) exemplificada anteriormente, para se mostrar o mecanismo de coerência no processo indexação-recuperação.

As informações do registro catalográfico são de autoria da BU.

Exemplo 1 : BOEHN, Max Von. La moda: historia del traje en europa.1945.

CATALOGAÇÃO PRÉ-MARC

Número de Chamada	391(4)(09) B671m
Autor Principal	Boehn, Max Von
Título Principal	Moda : Historia del traje en Europa(la) Barcelona : Salvat, 1945.
Publicação	 *Esta é a carta de indexação da obra, seguindo-se o procedimento descrito anteriormente: (de cima para baixo e da esquerda para a direita) Imagem icônica da obra, imagem indicial geográfica, e duas imagens assessórias para situar a época. O resultado é o conjunto chamado "carta de indexação". Abaixo, vêm-se os indexadores usados pela BU para descrever a obra.*
Assuntos	Adornos:antropologia cultural Aspecto historico de determinado assunto Elegancia:antropologia cultural Europa Fontes historicas Moda:antropologia cultural Trajes:antropologia cultural Vestuario:antropologia cultural

Exemplo 2: AMARO, Ana Maria. O Traje da mulher macaense : da saraça ao dó das nhonhonha de macau. 1989.

CATALOGAÇÃO PRÉ-MARC

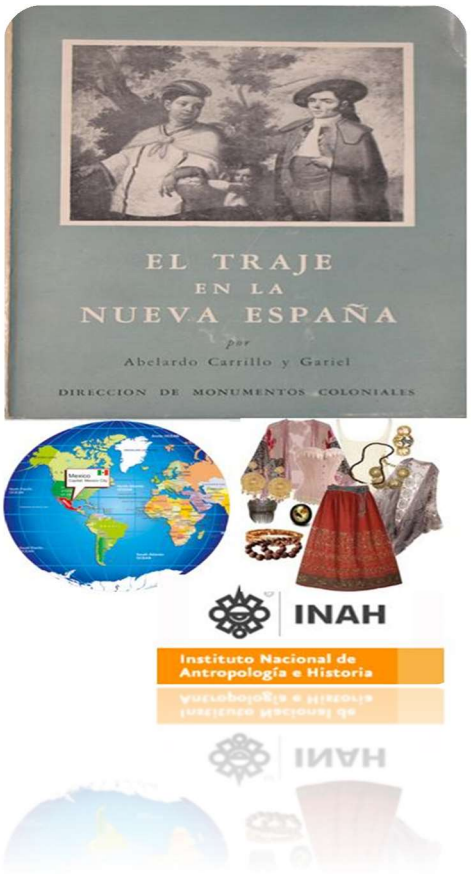
Número de Chamada	391.2(512.318) A485t
Autor Principal	Amaro, Ana Maria
Título Principal	Traje da mulher macaense : da saraca ao do das nhonhonha de macau(o)
Publicação	Macau : Instituto Cultural de Macau, 1989. 199 p. : il.
Descrição Física	*Nesse caso, a carta da obra foi composta com apenas 3 imagens auxiliares, devido à alta especialidade do tema, todavia completando a tríade ícone, índice e símbolo*
Assuntos	Condicoes da mulher:antropologia cultural Feminismo:antropologia cultural Macau:china:asia Moda feminina:antropologia cultural Mulher:antropologia cultural Trajes femininos:antropologia cultural Vestuario feminino:antropologia cultural

Exemplo 3 - CARRILLO Y GARRIEL, Abelardo. El traje en la nueva españa. 1959.

CATALOGAÇÃO PRÉ-MARC

Número de Chamada	391(72) C317t
Autor Principal	Carrillo Y Garriel, Abelardo
Título Principal	Traje en la Nueva Espana(el)
Publicação	México, DF : Inst Nac Antropol E Hist, 1959. 207 p. : il.

Descrição Física



Aqui tem-se a aplicação dos mesmos princípios, com a introdução de uma quarta imagem que situa o usuário na dimensão da autoridade da obra: o Instituto de antropologia e história. Observa-se ainda a translação do elemento indicial quando se lê na obra "Noueva España" que hoje é o México.

Assuntos

Adornos:antropologia cultural
Elegancia:antropologia cultural
Mexico:america do norte
Moda:antropologia cultural
Trajes:antropologia cultural
Vestuario:antropologia cultural

Considerações finais

A representação do conteúdo por meio de imagens está sustentada pelas teorias que fundamentam tanto semiologia saussuriana como a semiótica peirciana. A descrição das imagens como suas representações textuais mostram as bases para

os sistemas de gestão de bases de dados imagéticas, ferramentas essenciais para a composição do princípio de indexação imagética anunciado neste trabalho. As diretrizes revisitadas de Bertin mostram como se pode chegar à composição de imagem perfeitamente em acordo com as teorias linguísticas que explicam com clareza o processo de indexação textual e por conseguinte os mecanismos da indexação por imagem.

Os exemplos propostos para a navegação em OPACs imagéticos, assim como a composição das fichas descritivas dos documentos comprovam a aplicação das bases teóricas citadas e esclarecem sobre a viabilidade de uma indexação mais eficaz. Nesses termos, os resultados abrem possibilidades múltiplas para a acessibilidade aos acervos documentários. O público alvo se vê mais amplo, e maiores as oportunidades para o desenvolvimento dos hábitos de leitura. Os acervos imagéticos como a indústria editorial da imagem, e em particular as estórias em quadrinhos, descobrem novas formas de expansão nessa proposta. Os profissionais da informação desenvolvem novas habilidades com repercussões diretas na estrutura interna dos sistemas de gestão da informação, e na recuperação da informação igualmente. Da mesma forma, os sistemas de busca na web se enriquecerão dessas estratégias de indexação e recuperação da informação.

A metodologia proposta abre, assim, campo para novos desenvolvimentos de interfaces mais adaptados às comunidades com necessidades especiais, aos deficientes e aos iletrados funcionais. Espera-se a realização de estudos mais aprofundados sobre os mecanismos de percepção semântica das imagens. Competências transdisciplinares em neurociência, linguística e ciências da informação podem juntas explorar essas questões tais como a natureza dos processos cognitivos dos surdos em organização e classificação de informações, o funcionamento das gramáticas de comunicação dos autistas, os mecanismos de ação da biblioterapia para idosos, pacientes com Alzheimer e outras patologias depressivas, entre outras. Enfim, o uso da imagem para o estabelecimento de elos entre usuários e documentos parece encontrar nos sistemas de recuperação da informação por imagens um vasto campo de verificação científica tanto em pesquisa como em desenvolvimento.

Referências

ABNT. **Métodos para análise de documentos:** determinação de seus assuntos e seleção de termos de indexação. Rio de Janeiro: ABNT, 1992. 4p.

ASHTON, Danny.**10 reasons why you should care about visual content marketing.** Blog. Publicado em 19 de fevereiro, 2015. Disponível em: <http://neomam.com/blog/13reasons/>. Acesso em: 5 maio 2015.

BARTHES, R. **Le système de la mode**. Paris: Seuil, 1967.

BERTIN, Jacques. La graphique. **Communications**, v. 15, n. 1, p. 169–185, 1970.

BIBLIOTECA pública. Governo dos Açores. Disponível em: <http://www.bparah.azores.gov.pt/html/bparah-seccao+infanto-juvenil+cdu.html>. Acesso em: 3 maio 2015.

BOULANGER, Chantale. La surimage. **Applied Semiotics/Sémiotique appliquée**, v. 1, n. 2, p. 159–172, 1996.

BRITO, Marcilio de. Sistemas de informação em linguagem natural: em busca de uma indexação automática. **Ciência da Informação**, v. 21, n. 3, p. 223–232, 1992.

DANTIER, Bernard. La représentation et l'étude visuelles des informations. In.: BERTIN, Jacques. **Semiologie graphique, les diagrammes, les réseaux, les cartes**. 4.ed. Paris. Ehess, 2008. 32 p.

GHEORGHITA, Inga. Methodologie de construction automatique du thesaurus pour l'indexation et la recherche des images. **RECITAL**, n. 2, p. 221–228, Jun. 2011. Disponível em: <https://hal.archives-ouvertes.fr/hal-00695722>. Acesso em: 2 maio 2015.

GREIMAS, A.J. **La mode en 1830:** Essai de description du vocabulaire vestimentaire d'après les journaux de mode de l'époque. Paris, Sorbonne (thèse dactylographiée). s.d.

GREIMAS, A.J. **Sémantique Structurale recherche de méthode**. Paris: Larousse, Langue et Langage, 1966.

GREIMAS, A.J. Du Sens. **Essais Sémiotiques**. Paris: Seuil, 1970.

GREIMAS, A.J. Les Actants, les Acteurs et les Figures, In : **Sémiotique narrative et textuelle**, Claude Chabrol (dir.), Paris, Larousse, 1973, p. 161-176.

GREIMAS, A.J. **Sémiotique et Sciences Sociales**, Paris, Seuil, 1976. 219 p.

GUIRAUD, P. **De la grive au maquereau** : le champ morpho-syntaxique des noms de l'animal tacheté. Le français moderne. v.34, 1966

LE GUERN, M. **Sur les relations entre terminologie e lexique**. META Journal de traducteurs. Montréal. 1989

LEFEBVRE, Martin. The art of pointing: on Peirce, indexicality, and photographic images. **Photography Theory**, p. 220-244, 2007

MOUNIN, G. Essai sur la structuration du lexique de l'habitation. **Cahiers de lexicologie**, v.5, 1965.

MOUNIN, G. Un champ sémantique: la dénotation des animaux domestiques. **La linguistique**. v.1, 1965.

ROGERS. Simon. **What fuels a tweet's engagement?** (@smfrogers). Data editor 03/10/2014. Disponível em https://blog.twitter.com/2014/what-fuels-a-tweets-engagement Acesso em abril, 2015.

SOCIAL BAKERS. **Photos cluttering your facebook feed? here's why pictures account for majority of brand facebook posts**. Abril 21, 2014, Advertising & Marketing. Disponível em http://www.emarketer.com/Article/Photos-Cluttering-Your-Facebook-Feed-Herersquos-Why/1010777/1 Acesso em maio, 2015.

UNISIST. Princípios de indexação. **Revista da Escola de Biblioteconomia da UFMG,** Belo Horizonte, v. 10, n. 1: p. 83-94, mar. 1981.

ZIPF, George K. **Human behavior and the principle of least effort**. 1949.

Mapas conceituais interativos como instrumento didático no campo da Organização da Informação e do Conhecimento

Marilda Lopes Ginez de Lara
Universidade de São Paulo
larama@usp.br

Gabriela Previdello
Universidade de São Paulo
gprevidello.ecausp@gmail.com

Nair Yumiko Kobashi
Universidade de São Paulo
nairkobashi@gmail.com

1 Introdução

Ao lado das concepções que fundamentam o empreendimento de organizar a informação e o conhecimento para torná-los acessíveis aos usuários é fundamental investir também na formação de pessoas que realizem as atividades informacionais de forma consistente, sintonizada com as demandas dos diferentes segmentos sociais. É nesse espírito que desenvolvemos o projeto ora apresentado, na crença de que a formação deve ser emancipadora, substituindo o padrão baseado na transmissão linear e unidirecional do conhecimento, por vezes praticada na área. Acreditamos que o papel atual do professor é, antes de tudo, de tutoria e orientação, disponibilizando para o aluno um conjunto de conhecimentos para que ele realize criativamente a experiência da aprendizagem.

Adotando esses princípios e valendo-nos da possibilidade de acesso aos recursos tecnológicos oferecidos pela Universidade de São Paulo, no âmbito do Programa Pró-Ensino, desenvolvemos o projeto “Organização da informação e do conhecimento: mapas conceituais interativos na construção de informação

documentária". O projeto procura integrar as disciplinas da área de Organização da Informação e do Conhecimento do Depto de Biblioteconomia da Escola de Comunicações e Artes da Universidade de São Paulo – ECA/USP, a partir de princípios e procedimentos compartilhados. Pretendemos, com esta atividade, aprofundar a discussão sobre modos alternativos de formação acadêmica e profissional na área, fornecendo recursos didáticos no ambiente das redes de informação e comunicação.

O campo da Organização da Informação e do Conhecimento tem, entre seus objetivos, refletir criticamente e desenvolver as bases teóricas e metodológicas da organização, localização, recuperação e utilização do conhecimento registrado (CINTRA et al., 2002; HJORLAND, 2018). A denominação do campo, bem como a sua organização em disciplinas, varia muito nos cursos de graduação em Biblioteconomia em razão das vertentes teórico-metodológicas eleitas por cada instituição. Se a variação pode ser saudável, problemática é a formação do aluno, que nem sempre compreende o campo como conjunto integrado e interrelacionado de processos. A segmentação em diferentes disciplinas requer, portanto, um trabalho integrador orientado.

Nessa perspectiva, o projeto procura abordar o ensino-aprendizagem de forma aberta, sistematizando conteúdos que dizem respeito às teorias, métodos e processos de construção da informação documentária, acessível em plataformas digitais. Sua justificativa é a necessidade de reunir, para o aluno (e para o professor), um conjunto de referências de diferentes tipos e níveis de apresentação, de modo a complementar os estudos realizados presencialmente. Recorre, para isso, a instrumentos do próprio campo de Organização da Informação e do Conhecimento, permitindo o acesso a múltiplos conceitos e à identificação de suas relações.

2 Objetivos e justificativa

O objetivo geral do Projeto "Organização da informação e do conhecimento: mapas conceituais interativos na construção de informação documentária" é, em um primeiro momento, oferecer conhecimentos, em ambiente digital, aos alunos de Graduação do Curso de Biblioteconomia e Documentação da ECA/USP. Compreende os conteúdos das seguintes disciplinas: Introdução à Análise Documentária, Elementos de Lógica para a Documentação, Linguagens Documentárias I e II,

Linguística Documentária, Indexação: teoria e métodos. As disciplinas, oferecidas entre o 1º. e o 5º. semestres do curso, são ministradas pelos docentes que participam do projeto.

Concebido para ser desenvolvido em fases, o projeto deve ser gradativamente enriquecido com novos conteúdos e recursos, além de ser submetido a sucessivas avaliações.

Os objetivos específicos do projeto são:

- Desenvolver recursos de aprendizagem relativos aos tópicos principais das disciplinas, enfatizando seus conceitos, metodologias de trabalho, fluxos de tarefas, com destaque para as relações entre as operações documentárias;
- introduzir os conceitos da Análise Documentária enfocando os processos de condensação e representação de textos (verbais e imagéticos);
- introduzir os conceitos da Lógica, da Linguística e da Terminologia explorando as diferenças entre Linguagem Natural e Linguagem Documentária;
- apresentar as principais operações de indexação de documentos textuais e imagéticos, explorando os processos de representação documentária por meio de Linguagens Documentárias;
- discutir os aspectos estruturais das Linguagens Documentárias e as metodologias de controle de vocabulários;
- introduzir os princípios de promoção da interoperabilidade de vocabulários visando seu compartilhamento e reuso no ambiente da Web semântica;
- apresentar exemplos e sugerir atividades de aplicação das metodologias.

3 Princípios teórico-metodológicos

O pressuposto do projeto é o de que as atividades de Organização da Informação e do Conhecimento se desenvolvem, necessariamente, no universo da linguagem, razão pela qual suas referências teóricas baseiam-se, entre outras, nas Ciências da Linguagem, na Lógica, na Terminologia. O campo da Organização da Informação e do Conhecimento visa, como o próprio nome o diz, à organização dos objetos de conhecimento em sua singularidade (textos, imagens, tabelas estatísticas etc.), reunidos em sistemas, para colocá-los à disposição para acesso e utilização em suportes variados (arquivos, catálogos, bases de dados, sites), observando os diferentes níveis de granularidade dos dados (objetos individuais, objetos complexos

organizados em sistemas etc.), bem como as referências culturais e de linguagem do usuário. Do mesmo modo, visa refletir sobre tais práticas para compreender seus fundamentos, compromissos e posicionamento ético.

No âmbito dos estudos da linguagem são fundamentais os conceitos de Linguagem Natural e Linguagem Documentária (ou linguagem construída artificialmente), a noção de estrutura linguística (estrutura mínima de significação), de estruturação (sintaxe) e de uso (pragmática), entre outros. São também essenciais os conhecimentos sobre a natureza da linguagem, sua função simbólica, os fenômenos da enunciação, da discursividade e da textualização, bem como a distinção entre as operações de expansão e de condensação e a organização lógico-semântica dos vocabulários para representar conteúdos informacionais. Da Semiótica, são importantes os conceitos de objeto, signo, semiose, interpretante, contexto e intérprete, e da Terminologia, os conceitos de domínio do conhecimento ou área de atividade, formação e exploração de *corpora*, termo, conceito, relações entre conceitos, árvore de domínio, relações de equivalência conceitual e linguística e de validação social dos termos. Da Lógica, é essencial a noção de conceito, sua ordenação lógica, baseada na identificação da extensão e intensão das propriedades do conceito. Importantes também são as noções de equivalência conceitual e relações de espaço e tempo para operacionalizar e enriquecer a organização dos descritores em ferramentas de representação de informação para recuperação.

Enfim, o campo da Organização da Informação e do Conhecimento procura cuidar dos aspectos que dizem respeito à significação e às relações de sentido, reconhecendo as diferenças que caracterizam o discurso documentário dos discursos em linguagem natural ou especializada. Ao procurar representar os textos para os fins da recuperação, o campo recorre a operações que permitam assegurar relações de sentido por meio de oposições e contrastes, hiponímia ou subordinação, hiperonímia ou superordenação, deslocamentos genéricos e colaterais, relações parte-todo. Via Terminologia, o campo da Organização da Informação e do Conhecimento atualiza a ideia de campo semântico ao circunscrever seus vocabulários a domínios de especialidade e provê-los de estruturação a partir de referências pragmáticas.

A atitude interdisciplinar procura pautar a organização do campo: os conceitos não são meramente tomados de empréstimo de outros domínios; antes, são evocados para melhor compreender as manifestações linguístico-comunicacionais como

também para fundamentar as operações metodológicas do domínio específico da Organização da Informação e do Conhecimento.

Do ponto de vista didático-pedagógico, partimos do princípio que a construção do conhecimento exige articular o conceitual ao empírico. Com efeito, o campo não se restringe à observação e interpretação de fenômenos. Tem-se o compromisso com a criação de teorias e métodos para tratar informação para usos sociais. A organização da informação e do conhecimento encontra seu campo de validação, portanto, na eficácia com que a informação pode ser recuperada e utilizada socialmente.

As recentes discussões sobre a Web Semântica têm revitalizado as teorias de base da organização de conceitos. Investimentos em interoperabilidade realizados pela W3Consortium, pela IFLA e consolidados em normas internacionais de Documentação, como os padrões ISO (ISO 25964, 2011; 2012), têm recolocado a importância do conceito e dos sistemas conceituais como bases para a representação e a organização de vocabulários controlados.

4 Método

Os conceitos de ORC apresentados nesta fase do projeto foram selecionados segundo sua maior importância para as disciplinas do Curso de Biblioteconomia da ECA-USP. Os conteúdos programáticos das disciplinas e sua bibliografia constituíram o corpus de identificação e seleção dos conceitos. Para que as relações entre os conceitos tenham maior visibilidade, fato nem sempre percebido pelos alunos, foram eles organizados em categorias temáticas e não por disciplinas. Buscou-se compreender tais categorias num sentido pragmático: agrupamentos que manifestam particularidades em relação a uma área de conhecimento (no nosso caso a Organização da Informação e do Conhecimento) que servem a propósitos institucionais ou pontos de vista particulares, não sendo, portanto, de aplicação universal.

Operacionalmente, os conceitos selecionados foram, num primeiro momento, organizados sob a forma de tesauro, uma vez que a metodologia de construção dessas linguagens mostrou ser adequada para abrigar e distribuir estruturalmente os termos. Para armazenar os termos que correspondem a conceitos recorreu-se primeiramente ao Multites (Multisystems, 2005-2014) e, em seguida, ao TemaTres

(TemaTres, 2015), considerando não apenas o fato de este último ser um software aberto e gratuito, com a vantagem de ser desenhado para promover a interoperabilidade entre diferentes vocabulários controlados.

O projeto está estruturado em módulos que apresentam os seguintes conteúdos básicos:

- aulas em vídeo com duração de 5 a 10 minutos;
- literatura específica: arquivos com textos fundamentais;
- referências, fontes de informação;
- exemplos de aplicação;
- endereços de sites com conteúdos relacionados;
- softwares relacionados;
- exercícios.

Os conteúdos são estruturados sob a forma de mapa conceitual, com a diferença de que seu desenho apóia-se na hierarquia, na associação e na equivalência para garantir maior consistência à rede de relacionamentos, tal como prevê a metodologia de organização de um tesauro. Os principais nós do mapa correspondem a categorias temáticas que, por sua vez, compreendem conceitos e seus relacionamentos com outros conceitos dentro da própria categoria, ou com conceitos de outras categorias. Os conceitos centrais são, por sua vez, *linkados* a referências bibliográficas, exemplos, exercícios etc.

O suporte tecnológico utilizado compreende um sistema de gerenciamento de conteúdo para site CMS em Wordpress, com layout nas linguagens HTML4.0 Transational/ CSS/PHP 5.2/ JS, e banco de dados MySQL 5.1, além de outros padrões de desenvolvimento como o JQuery versão 1.5x e Javascript versão 1.1x. Os requisitos tecnológicos contemplados foram os de usabilidade, portabilidade e interoperabilidade, a fim de que as linguagens fossem amplamente compreendidas e compatíveis com todos os navegadores. O sistema opera de forma integrada às ferramentas em uso na Universidade de São Paulo, a exemplo da plataforma Moodle – STOA – USP, e a softwares especializados relacionados às disciplinas (desenvolvimento de vocabulários controlados, mapas conceituais, visualização de informações) aos serviços de mídia (Youtube, Vimeo, Soundcloud) e às mídias sociais (Facebook, Google+, Twitter).

O projeto gráfico procurou equacionar a proposta de ensino/aprendizagem utilizando formas alternativas de visualização dos conteúdos agrupando, em um mesmo espaço, os conceitos das várias disciplinas em categorias temáticas utilizando cores sólidas para destacá-las. A implementação do projeto partiu de 200 termos teóricos e operacionais selecionados do conteúdos programáticos e bibliografia básica das disciplinas. Os termos foram então organizados em uma estrutura padronizada, que integra, para cada termo, além de sua rede de relações, a definição conceitual ou características do conceito, famílias de termos, exemplos de uso, exercícios, videos etc. A forma de estruturação adotada permite que o aluno, ao escolher um termo ou uma categoria temática, descubra as múltiplas relações que existem entre os conceitos teóricos e operacionais, bem como as informações correspondentes que os contextualizam.

As informações estão distribuídas em um painel que permite a atualização dos dados de forma prática e intuitiva. A leitura pública dos dados pode ser feita através de uma API (http://www2.eca.usp.br/oc/sobre-api/), de modo a, num futuro próximo, disponibilizar as informações a outras instituições e usuários.

Para as fases seguintes, estão previstos espaços para conferências e discussões virtuais entre alunos e professores, bem como para acolhimento de críticas, observações, exposição de dificuldades e sugestões. De forma eventual, os módulos deverão conter, também, filmes, vídeos com relatos de experiência de outras instituições, propostas de soluções para comparação, passeios virtuais.

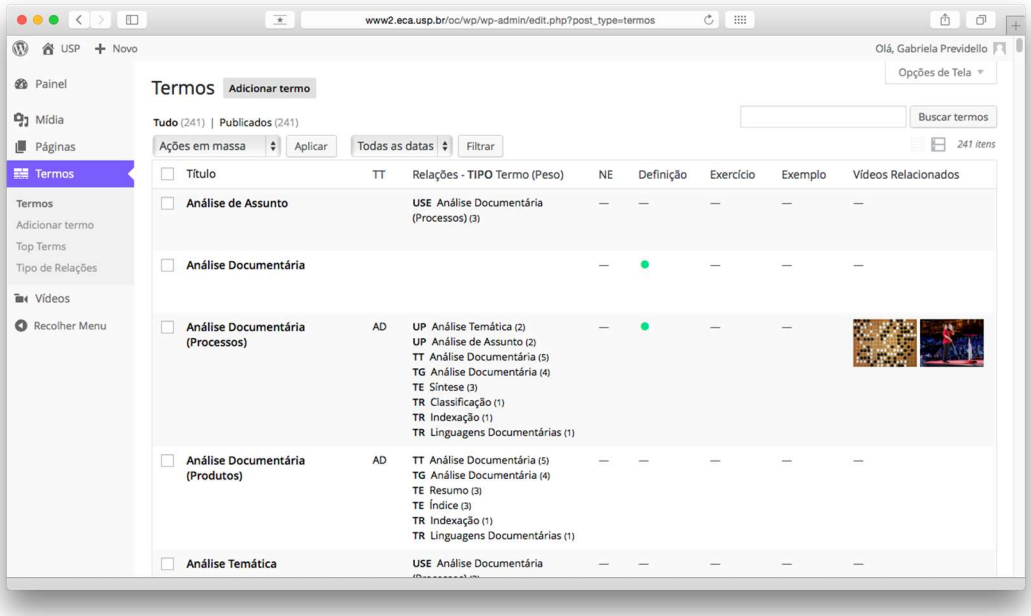
5 Resultados

Apresentamos, neste item, a plataforma de trabalho e o resultado da organização preliminar de dados e visualizações parciais de categorias e termos na forma de mapas conceituais.

5.1 Plataforma de trabalho

A plataforma de trabalho apresenta as seguintes funcionalidades:

Figura 1 – Tela de trabalho: inclusão de termos e suas relações



5.2 Organização dos termos

Os conteúdos da organização preliminar dos dados podem ser visualizados a seguir:

Figura 2 – Visualização alfabética de termos

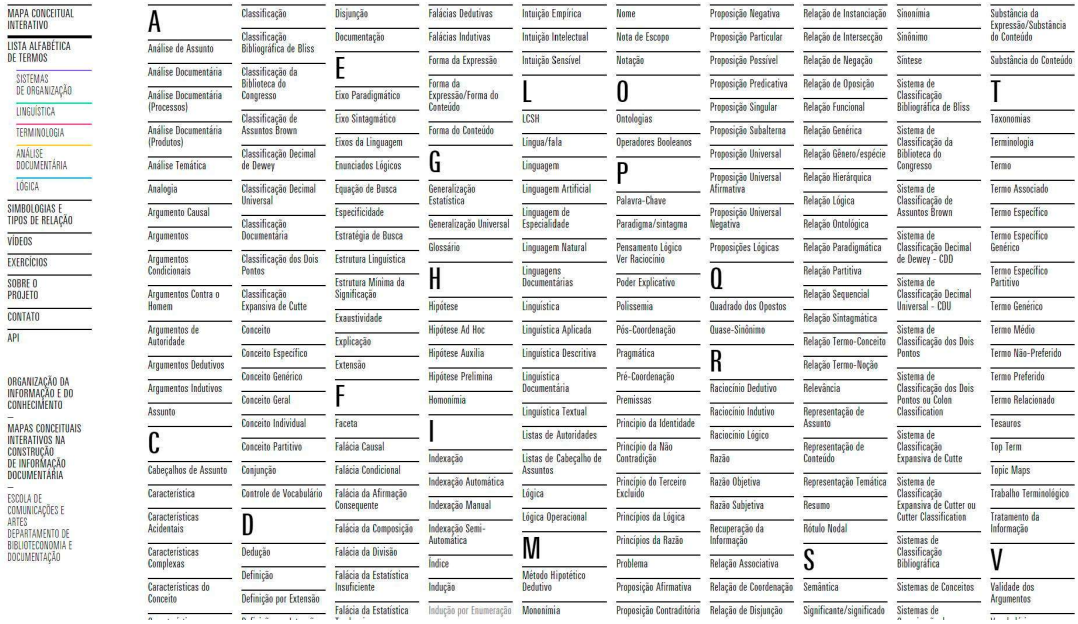


Figura 3 – Visualização gráfica da categoria “Sistemas de organização do conhecimento”

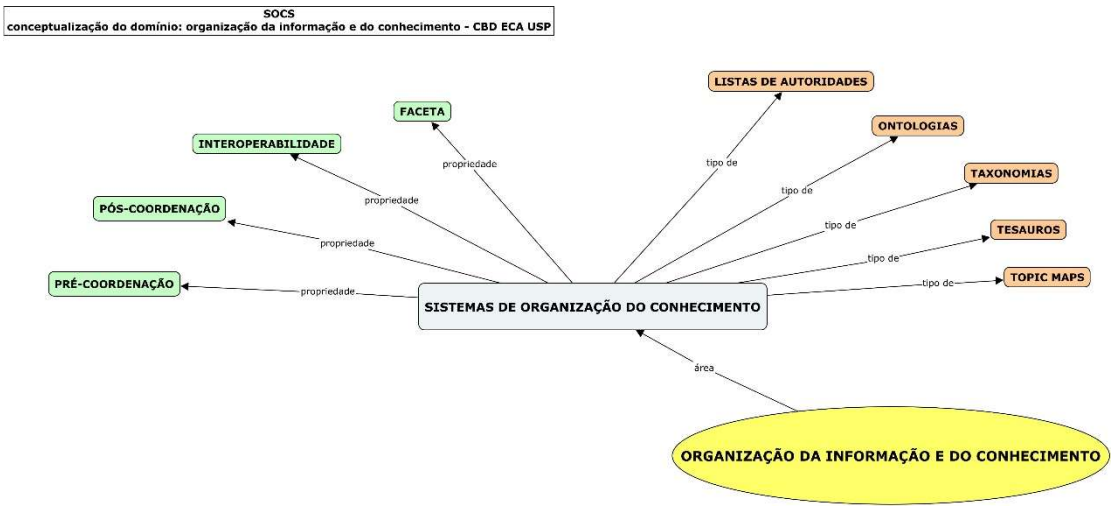
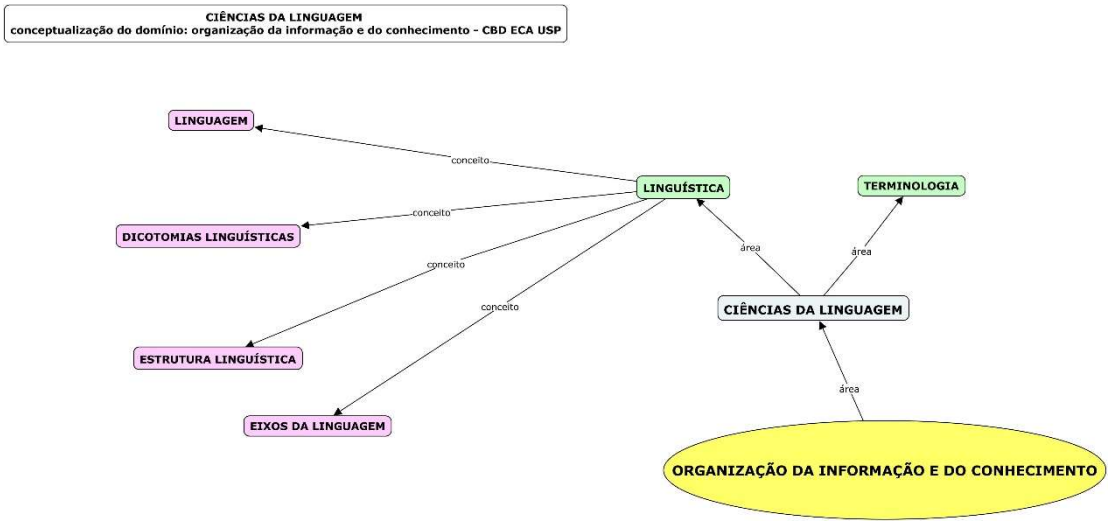


Figura 4 – Visualização gráfica da categoria “Ciências da linguagem”



O desenvolvimento do projeto tem mostrado que, além dos benefícios para os alunos de graduação, ele desencadeia, entre os docentes, um processo de discussões produtivas sobre o plano didático-pedagógico, principalmente em relação aos conteúdos privilegiados e à integração entre as disciplinas. Do mesmo modo, o projeto coloca em evidência a necessidade de investir na produção de um esquema de equivalências entre conceitos que têm origem nas diferentes teorias que convivem na área, como também nos fundamentos das escolhas realizadas.

Outro aspecto importante é a validação que deve ocorrer em 2 semestres contínuos, previstos para o 2º. Semestre de 2015 e 1º. Semestre de 2016, período mínimo para garantir que o aluno frequente ao menos 2 disciplinas e possa interagir com os mapas conceituais disponíveis na plataforma como meio de complementar sua formação.

6 Considerações finais

A renovação do ensino de Organização da Informação e do Conhecimento enfrenta inúmeros desafios: a concepção de um modo de organizar conteúdos, com fins didáticos; a determinação dos conceitos nucleares e periféricos e suas formas de interação; as competências necessárias para a implantação da proposta; a escolha das tecnologias apropriadas para estes fins e, não menos importante, a captação de recursos para as atividades de ensino, sabidamente mais escassos do que os aportes financeiros disponíveis para a pesquisa.

A escolha do eixo da proposta - a categorização de conceitos – foi definida porque permite mostrar de forma integrada, articulada, os principais conceitos e operações da Organização da Informação e do Conhecimento. Procurou-se, com esta abordagem, tentar superar a visão fragmentada do campo. De fato, se a disciplinarização, ou recorte do conhecimento em disciplinas permite aprofundar conhecimentos, pode, por outro lado, promover a dispersão dos conceitos e operações de um dado domínio. É necessário, portanto, na formação, investir em ações que promovam um olhar integrador. O uso de diferentes visualizações gráficas, como exemplificado acima, é, certamente, um diferencial importante desta proposta.

Para a exploração dos recursos oferecidos pelo ambiente digital foram adotadas diversas formas de apresentação dos conteúdos: aulas em vídeo; *links* para textos fundamentais; fontes de informação; exemplos de aplicação, sites com conteúdos relacionados; *softwares* relacionados; exercícios.

O suporte tecnológico utilizado compreende um sistema de gerenciamento de conteúdo para *sites,* linguagens apropriadas, banco de dados relacionais. Os requisitos tecnológicos foram selecionados com base nos conceitos de usabilidade, portabilidade e interoperabilidade. Finalmente, o sistema opera de forma integrada com as ferramentas em uso na Universidade de São Paulo, a exemplo da plataforma Moodle – STOA – USP e redes sociais.

Espera-se, ainda, que os resultados deste projeto possam ser utilizados pela comunidade que se dedica ao ensino de Organização da Informação, matéria indispensável para a formação de bibliotecários e demais profissionais da informação.

Agradecimentos

Agradecemos à equipe atual do projeto, em especial às Profas. Dra. Cibele Marques dos Santos, Dra. Giovana Deliberali Maimone, Dra. Vânia M. Alves Lima, e à estagiária (aluna de Graduação) Lúcia Marques. Nossos agradecimentos, também, ao Programa Pró-ensino da USP, pelos recursos concedidos para o desenvolvimento do projeto.

Referências

CINTRA, A.M.M.; TÁLAMO, M.F.G.M.; LARA, M.L.G.; KOBASHI, N.Y. **Para entender as linguagens documentárias**. São Paulo : Polis, 2002.

HJORLAND, B. Knowledge Organization Systems (KOS). In: HJORLAND, B. **Lifeboat of Knowledge Organization**. 2008. Disponível em:

<http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm>
Acesso em mar. 2015.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 25964:** thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. Geneve: International Standard Organization, 2011.

INTERNATIONAL STANDARD ORGANIZATION. **ISO 25964:** thesauri and interoperability with other vocabularies. Part 2: Interoperability with other vocabularies. Geneve: International Standard Organization, 2012.

MULTISYSTEMS, 2005-2014. Disponível em: http://www.multites.com/>
Acesso em jul. 2015.

TEMATRES, 2015. Disponível em: http://www.vocabularyserver.com/index.html Acesso em jul. 2015.

Análise de domínio na organização do conhecimento: explorando as relações temáticas e de citação

Bruno Henrique Alves
Universidade Estadual Paulista
brhenriquealves@gmail.com

Ely Francina Tannuri de Oliveira
Universidade Estadual Paulista
etannuri@gmail.com

Maria Cláudia Cabrini Grácio
Universidade Estadual Paulista
cabrini@marilia.unesp.br

Introdução

A Organização do Conhecimento é uma disciplina que trata do desenvolvimento de técnicas para construção, gestão e uso, avaliação de classificações científicas, taxonomias, nomenclatura e linguagens documentais. Trata, ainda, de metodologias de uso e recuperação por linguagem natural (NAVARRO; MARCO, 1995, BARITÉ, 2001). Desta forma, o objeto de estudo da Organização do Conhecimento, segundo os autores em questão, é o conhecimento que, socializado, possa gerar novos conhecimentos.

Atualmente, o estudo de seu domínio tem sido alvo de vários estudos, tais como: Guimarães; Santos; Cândido; Pinho (2014), Smiraglia (2014) e Arboit (2012), entre outros, que se propõem a definir e conceituar seu escopo, seu objeto de estudo, seus fundamentos teóricos e, principalmente, delinear metodologias que melhor

possam compreender o conhecimento construído.

Hjørland e Albrechtsen (1995) desenvolveram um paradigma social epistemológico denominado *Análise de Domínio* (A.D.), cuja vertente principal destaca a relação existente entre os diferentes campos do conhecimento e as comunidades discursivas. É um processo pelo qual se pode verificar o que é significativo em alguma comunidade científica, apontando elementos que permitem analisar seu contexto científico, tais como tendências, padrões, objetos (pessoas), processos e relações existentes, envolvendo a compreensão do cultural e antropológico, do histórico e social.

Nesta pesquisa, o domínio é a área de conhecimento denominada O.C., em que se demarcam os limites, com profissionais ou grupos articulados em pensamento e em linguagem.

Entre as 11 abordagens de análise de domínio apresentadas por Hjørland (2002), destacam-se os estudos bibliométricos, que, associados a estudos epistemológicos ou históricos, produzem uma interação fértil e melhor visualização das diferentes áreas científicas, revelando correntes teóricas dominantes, seja por meio de colaborações científicas, seja por meio das análises de citações ou de cocitações.

O conjunto de referências bibliográficas utilizadas na elaboração de um documento pode ser analisado como reflexo de uma comunidade discursiva composta por pesquisadores e suas temáticas, de modo a constituir um domínio científico.

Leydesdorff (1998) observa que as citações apontam tanto o contexto social quanto cognitivo do conhecimento nas práticas científicas, sustentam a comunicação científica e definem a estrutura intelectual de uma área do conhecimento.

A análise de citação consiste em uma forma de definir um domínio, ao identificar os autores de maior inserção e impacto, o diálogo e o correconhecimento estabelecido entre eles pela comunidade citante (SMIRAGLIA, 2011; GLÄNZEL, 2003). Aponta seus paradigmas, procedimentos metodológicos pertinentes e os pesquisadores de “vanguarda”.

Considerando o que foi exposto, esta pesquisa objetiva analisar os pesquisadores que contribuem para o desenvolvimento da temática Organização do Conhecimento (O.C.) por meio da análise de domínio, em sua abordagem de estudos bibliométricos.

Especificamente, objetiva identificar e apresentar, de forma diacrônica, a elite científica em O.C., caracterizada pelos pesquisadores mais produtivos na revista *Knowledge Organization*. Ainda, destacar as temáticas contempladas pelos pesquisadores mais produtivos e evidenciar seus referentes teóricos e metodológicos, bem como a rede de interlocução desta elite com os autores mais citados, nos quinquênios 2004-2008 e 2009-2013.

Esta pesquisa justifica-se pela sua contribuição analítica para a compreensão do domínio da Organização do conhecimento, ao oferecer subsídios teóricos e metodológicos relativos à elite científica e respectivos referentes.

Metodologia

Localizaram-se no periódico *Knowledge Organization,* indexado na base de dados *Scopus,* no período de 2004 a 2008, 41 artigos publicados por 57 autores e 74 artigos publicados por 105 autores, no quinquênio 2009-2013. Foram selecionados os 19 autores mais produtivos, correspondendo à publicação de pelo menos 2 artigos, considerando o período de 10 anos. Em seguida, separaram-se as publicações destes 19 pesquisadores pelos 2 quinquênios, para uma análise diacrônica mais detalhada em relação às citações.

Em vista das palavras-chave, agruparam-se as temáticas presentes nos 36 artigos dos 19 pesquisadores, utilizando-se o Sistema de Classificação da literatura em O. C., de Dahlberg, a fim de identificar as proximidades temáticas entre os autores mais produtivos.

Os autores foram agrupados segundo a distribuição da produção por quinquênio e classificação das temáticas, segundo o sistema de Dahlberg, por meio da análise de clusters, utilizando o método Ward e distância euclidiana. A visualização dos agrupamentos de pesquisadores está apresentada na Tabela 1, juntamente com a produção por quinquênio, total e distribuição dos artigos por categoria, de Dahlberg, de forma não excludente.

No primeiro quinquênio (2004-2008), para os 19 pesquisadores mais produtivos, encontraram-se 651 citações, relativas a 511 autores citados; e no segundo quinquênio (2009-2013), 775 citações, relativas a 780 autores citados. Foram considerados autores mais citados aqueles que receberam pelo menos 4 citações, totalizando 18 autores no primeiro quinquênio e 10 pesquisadores no segundo. As

autocitações foram eliminadas, e os pesquisadores que foram contemplados com mais de uma referência no mesmo artigo foram registrados apenas uma vez.

Por fim, construíram-se, no *software Excel*, as duas matrizes assimétricas 19x18, e 19x10, relativas aos mais produtivos e aos mais citados, uma para cada quinquênio. Utilizou-se o *software Ucinet*, a fim de gerar as redes *two mode* (citantes e citados).

Apresentação e análise dos resultados

A Tabela 1 apresenta os 19 pesquisadores mais produtivos, por quinquênio, e temática trabalhada, agrupados segundo proximidade pela análise multivariada de Cluster.

Tabela 1. Pesquisadores mais produtivos no periódico *Knowledge Organization*, por quinquênio, total e segundo Sistema de Classificação da literatura em O. C., de Dahlberg.

Pesquisador		2004 a 2008	2009 a 2013	Total	C1	C2	C3	C4	C5	C6	C7	C8	C9
Gnoli, C.	G1	3	1	4	3	1	2					1	
Kipp, M.E.I.		0	3	3	1	2	1				2		
Campbell, D.G.		0	2	2		2	1				1		
Mazzocchi, F.	G2	1	1	2		2	1				1		
Osińska, V.		0	2	2		1	1	1					
Olson, H.A.		1	2	3		2	2	1			1		
Dahlberg, I.		1	1	2	2	1							
Tennis, J.T.	G3	1	1	2	1		1						
Buchel, O.		0	2	2	1						1		
Mai, J.-E.		0	2	2		1					1		1
Golub, K.		1	1	2		1	2				1		
Lee, H.-L.	G4	1	1	2		1	2				1		
Souza, R.		1	1	2			2				1		
López-Huertas, M.J.		2	0	2		1					1		1
Ménard, E.	G5	1	1	2			1				2		
Park, J.-R.		1	1	2							2		
Chaudhry, A.S.		2	1	3		3			1				
Khoo, C.	G6	1	1	2		2							
Wang, Z.		1	1	2		2							

G1: total de artigos destacado e produção em 4 categorias distintas;

G2: todos pesquisadores publicaram no 2º quinquenio e em C2 e C3;

G3: presença nos 2 quinquenios e disperso nas categorias;

G4: todos pesquisadores publicaram nos 2 quinquenios e em C3 e C7;

G5: todos pesquisadores com artigos em C7 e pouca presença nas demais;

G6: todos pesquisadores com artigos nos 2 quinquenios e centrada em C2

Fonte: Elaboração dos autores

Observa-se que cinco pesquisadores não tiveram produção no primeiro quinquênio, e no segundo quinquênio somente um pesquisador não apresentou produção, significando assim uma ampliação e maior dispersão da elite produtora neste quinquênio.

Ainda, dos 19 pesquisadores, 10 deles publicaram um artigo no primeiro quinquênio e um artigo no segundo quinquênio, evidenciando uma relativa estabilidade de pesquisa na área.

Fazendo-se uma análise diacrônica da dinâmica da produção científica, nestes dez anos, observa-se que os cinco pesquisadores ingressantes no segundo quinquênio − Kipp, Buchel, Campbell, Mai e Osinska − são aqueles com maior produção, juntamente com Olson, que já compunha o grupo no primeiro quinquênio. Por outro lado, os pesquisadores mais produtivos no primeiro quinquênio diminuíram significativamente sua produção no segundo quinquênio.

Além disso, observa-se que os grandes produtores da O.C. e que publicaram na K.O. neste período são advindos dos E.U.A., do Canadá e da Europa, com destaque para um único brasileiro – Souza − da Fundação Getúlio Vargas.

Em relação às temáticas que mais se destacaram nos artigos, a partir da classificação usada, a mais contemplada é C2, relativa aos **Sistemas de Classificação e Tesauros, estrutura e construção,** que envolvem relações entre conceitos, taxonomias e linguagem de indexação, entre outros itens. A seguir, aparece também C3, relativo à **Metodologia de Classificação e Indexação,** incluindo teoria de indexação e classificação, análise de assunto, indexação automática e geração de índices e programas, e também C7, **Representação do Conhecimento por Linguagens e Terminologia,** que aborda linguagem natural em relação à O.C., semântica, sistemas de recuperação on-line, tecnologias e terminologias. Por outro lado, observa-se a inexistência de trabalhos em C6, referentes às **Classificações e Tesauros de assuntos específicos,** e somente um artigo de Chaudhry em C5, **Classificação orientados à objetos (Taxonomias), nas diferentes áreas de Conhecimento, e** C8, **Indexação e Classificação aplicadas.**

A partir desse relato de temáticas, compreende-se uma maior preocupação dos pesquisadores em O.C. de se voltarem para o próprio objeto da área, aprofundando os estudos de forma a consolidar os temas fundamentais da área, deixando menos

priorizados temas aplicados em outras áreas de conhecimento.

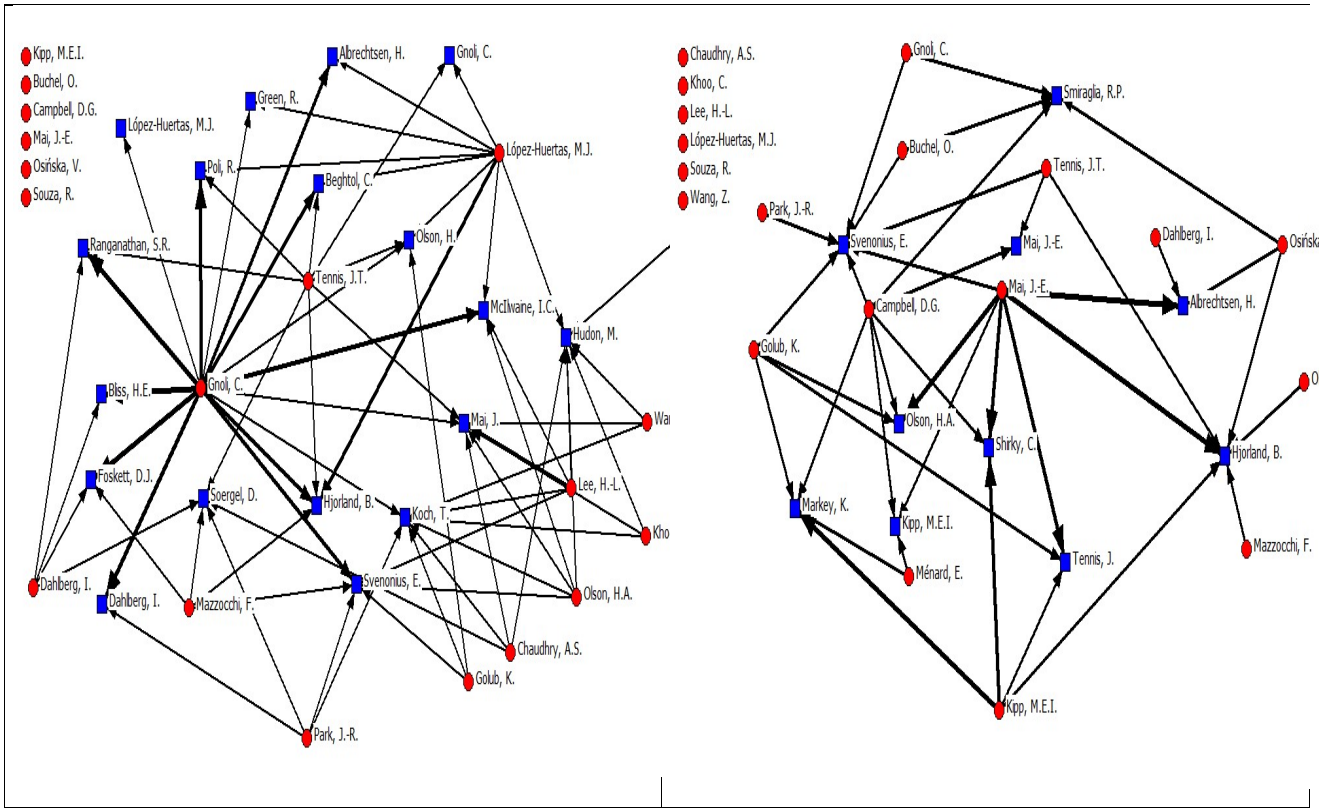
Apresentam-se, nas Figuras 1a e 1b, as duas redes de pesquisadores citante-citado, em que os círculos vermelhos são os citantes. A espessura dos segmentos de reta representa as frequências de citações. Observe-se que há cinco autores que aparecem citados nos dois quinquênios, a saber, Svenonius, Albrechtsen, Hjørland, Mai e Olson, pesquisadores de grande destaque nos estudos dos fundamentos da Organização do Conhecimento, no século XX .

Svenonius centra seus estudos na metaconstrução para a teoria da Organização do Conhecimento, tendo como trabalho seminal o livro *Intellectual Foundations of Information Organization.* Albrechtsen foca seus estudos nas seguintes áreas de pesquisa: gestão do conhecimento, análise de domínio, empreendedorismo. Hjørland é especializado em arquitetura da informação, literacia da informação, recuperação da informação e comportamento informacional. Mai estuda questões sobre construções conceituais, metodológicas e programáticas, assim como a natureza da classificação. A pesquisa de Olson centra-se na análise crítica de representações de assunto e sistemas de classificação, com perspectiva de caráter feminista, pós-estrutural e pós-colonial.

Na Figura 1a, aparecem 18 pesquisadores mais citados, com pelo menos 4 citações. Os pesquisadores Gnoli, Olson e Dahlberg, com, respectivamente, três, uma e uma publicações, pertencem também ao conjunto dos mais citados, com cinco, quatro e quatro citações no período. Ainda, observa-se nesta figura uma grande quantidade de ligações citante-citado, o que sugere a pluralidade e dispersão de referentes teóricos adotados no primeiro quinquênio.

Na Figura 1b, aparecem 10 pesquisadores mais citados, indicando que, nesta rede, o número de autores mais citados é menor que aquele presente na Figura 1a, todavia com maior concentração de frequência de citações, observada pelos segmentos mais fortes. Esse comportamento aponta que no segundo quinquênio há um grupo de pesquisadores mais consignados na área, evidenciando um fortalecimento das questões teóricas, com fundamentos mais determinados. Observa-se que Kipp, Mai, Olson e Tennis pertencem ao grupo dos mais produtivos e do mais citados, com quatro, seis, cinco e seis citações, respectivamente.

Figura 1a. Rede citante-citado (2004-2008)	Figura 1b. Rede citante-citado (2009-2013)



Fonte: Elaboração dos autores

Destaca-se, ainda, que, em cada quinquênio, seis dos autores mais produtivos não utilizaram em suas pesquisas, como referenciais teóricos, os autores mais citados: no 1º quinquênio, Kipp, Buchel, Campbell, Mai, Osińska e Souza; e no 2º quinquênio, Chaudhry, Khoo, Lee, López-Huertas, Wang e Souza, são pesquisadores que utilizam referentes teóricos distintos dos 13 pesquisadores restantes em cada um dos quinquênios. Assim, Souza, o único pesquisador brasileiro, em nenhum dos dois quinquênios desenvolveu suas pesquisas a partir dos autores mais citados.

Ainda, destacam-se os pesquisadores citados em maior número de artigos: no quinquênio 2004-2008, Koch, citado em 8 artigos, Mai em 6 artigos, Svenonius e Soergel em 5 artigos; e no quinquênio 2009-2013, Svenonius em 7 artigos, Hjørland citado em 6 artigos e Smiraglia e Markey em 4 artigos.

Destaque-se que, dos 19 pesquisadores mais produtivos, 6 são também os mais citados (~37%), a saber: Gnoli, Olson, Dahlberg, Kipp, Mai e Tennis, constituindo-se em um grupo que consigna a elite e a frente de pesquisa.

Conclusões

A pesquisa mostrou os pesquisadores, suas temáticas e referentes teórico-

metodológicos utilizados pelo periódico K.O., durante o período estudado, analisando o domínio em O.C., por meio da abordagem bibliométrica.

As principais temáticas pesquisadas privilegiam categorias de assunto mais nucleares para o desenvolvimento e consolidação da O.C.

Observe-se que Svenonius, Albrechtsen, Hjørland, Mai e Olson foram citados nos dois quinquênios, sinalizando que pesquisadores clássicos dos fundamentos da O.C. pertencem à frente de pesquisa da elite científica do periódico analisado. Ainda, Gnoli, Olson e Dahlberg pertencem tanto à frente de pesquisa como à elite deste periódico, por estarem entre os mais produtivos e mais citados.

Finalizando, destaca-se a maior pluralidade e dispersão de referentes teóricos adotados no primeiro quinquênio (2004 a 2008) em relação ao quinquênio seguinte, em que há um grupo de pesquisadores mais consignados na área, indicando fortalecimento das questões teóricas, com fundamentos mais determinados.

Referências

ARBOIT, A.E. O processo de (re)construção da teoria do conceito no domínio de Organização do Conhecimento: uma visão dialógica. **Scire**, v. 18, n. 2, p.129-134, 2012.

BARITÉ, M. Organización del Conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, Kester. (Org.). **Educação, universidade e pesquisa.** Marília, 2001.

ESTEBAN NAVARRO, M. A.; GARCIA MARCO, F. J. Las “Primeras Jornadas sobre Organización del Conocimiento: Organización del Conocimiento e Información Científica”, **Scire**, v.1, n.1, p.149-157, 1995.

GLÄNZEL, W. **Bibliometrics as a research field**: a course on theory and application of bibliometric indicators. Bélgica, 2003.

GUIMARÃES, J. A. C. DOS SANTOS, A. Y., CÂNDIDO, G. G., PINHO, F. A. A dimensão temática da pesquisa em organização do conhecimento: uma análise de domínio dos congressos nacionais e regionais da ISKO. **Scire**, v. 20, n. 2, p.19-25, 2014.

HJØRLAND, B.; ALBRECHTSEN, H. Toward a new horizon in Information Science: domain-analysis. **Journal of the American Society for Information Science**, v. 6, n. 6, p. 400-425, 1995.

HJØRLAND, B. Domain analysis in information science: eleven approaches traditional as well as innovative**. Journal of Documentation**, v. 58, n. 4, p.422-462, 2002

LEYDESDORFF, L. Theories of citation? **Scientometrics**, v. 43, n. 1, 1998.
SMIRAGLIA, R.P. II congresso brasileiro em representação e organização do conhecimento: Knowledge Organization in Rio 2013 - An editorial. **Knowledge Organization**, v.41, n.2, p.105-112, 2014.
SMIRAGLIA, R.P. ISKO 11’s Diverse Bookshelf: an editorial. **Knowledge Organization**, v. 38, n.3, p. 179-186, 2011.

Um estudo do domínio em organização do conhecimento por meio da análise de citação e cocitação de autores

Lidyane Silva Lima
Universidade Estadual Paulista
lidyane_lima17@hotmail.com

Pollyana Ágata Gomes da Rocha Custódio
Universidade Estadual Paulista
pollyanaagata@hotmail.com

Ely Francina Tannuri de Oliveira
Universidade Estadual Paulista
etannuri@gmail.com

Leilah Santiago Bufrem
Universidade Estadual Paulista
santiagobufrem@gmail.com

Introdução

Os estudos de Organização do Conhecimento (OC) têm papel de destaque no campo da Ciência da Informação desde seu reconhecimento, como ciência institucionalizada e vêm ensejando, historicamente, o fortalecimento dos marcos conceituais da Ciência da Informação e da Documentação, na medida em que fornecem as bases teórico-metodológicas para melhor compreender o processo nuclear dessa disciplina: a mediação entre o conhecimento produzido e a socialização desse conhecimento.

As pesquisas da área apontam para, no mínimo, dois aspectos importantes: de um lado, a abordagem teórico-filosófica voltada para questões epistemológicas ao analisar um domínio juntamente com as implicações na organização de sistemas de classificação; e, de outro lado, os métodos empregados na elaboração propriamente dita dos sistemas na gestão do conhecimento (taxonomia, tesauros e ontologias).

Considerando o exposto, esta pesquisa objetiva analisar por meio da análise de domínio, em sua abordagem de estudos bibliométricos, mais especialmente, análise de citação e cocitação, o contexto científico da área de "Organização do Conhecimento" (OC), a partir das publicações indexadas no periódico *Knowledge Organization (KO),* bem como determinar as grandes linhas que definem sua comunidade científica, em seus referentes teóricos e metodológicos.

Este estudo se justifica, pela necessidade de ampliação e aprofundamento das pesquisas em OC, para que se possa melhor sinalizar o estágio atual da ciência e do desenvolvimento da área de "Organização do Conhecimento", no contexto de abrangência internacional como o peródico KO (*Knowledge Organization).*

Análise de citação e cocitação de autores (ACA)

A citação pode ser entendida como o conjunto de uma ou mais referências bibliográficas de uma publicação, que evidencia as relações de uma publicação com outra. A análise de citação permite que se conheçam os autores mais influentes de um domínio científico por meio da frequência de citações registradas.

A análise de cocitação derivada da análise de citação trata da frequência com que dois documentos, autores ou periódicos são citados juntos em uma literatura e contribui para a compreensão da estrutura intelectual das ciências, evidenciando as relações e similaridades cognitivas e sociais entre estes autores, estabelecidas pela comunidade citante do domínio estudado.

Segundo White e McCain (1998), estudos relativos à Análise de Cocitação de Autores (ACA) tiveram como precursores White e Griffith, em 1981, e sua principal função é identificar os autores influentes e mostrar suas inter-relações, a partir das citações registradas. Em geral, pesquisadores com problemas de pesquisa semelhantes citam fontes informacionais similares.

A partir da construção das redes da ACA, é possível: (1) a identificação do grupo de autores mais cocitados; (2) a localização desse grupo de autores em suas

respectivas escolas de pesquisa e linha de pensamento; (3) a definição do grau de centralidade dos autores ou os autores centrais e os autores concebidos como periféricos dentro do conjunto desses mesmos autores; (4) a proximidade entre os autores e suas inter-relações e o grupo de autores que aparecem interligados, interfaciando diversas áreas de pesquisa e (5) visualizar a posição estrutural dos autores na rede de cocitação (WHITE; GRIFFITH, 1981, p. 163).

Nesse contexto, esta pesquisa busca contribuir com os estudos de Análise de Cocitação de Autores (ACA) ao analisar e destacar os pesquisadores de maior visibilidade e suas inter-relações na temática Organização do Conhecimento. Considera-se também que estudos de mapeamento de redes científicas devem ser desenvolvidos em associação com análises epistemológicas, críticas e históricas, a fim de se compreender um domínio de forma mais plena e aprofundada e se obter uma análise de caráter qualitativo.

Procedimentos metodológicos

Realizou-se uma busca na base *Scopus* pelo título da revista em estudo "*Knowledge Organization*", filtrado no campo "*source Title*". Depois foram recuperados todos os artigos publicados na revista em questão no período de 2004 a 2013, na modalidade artigo. A seleção do *corpus* de pesquisa foi realizada em novembro de 2014 e resultou em um total de 115 artigos. O mecanismo de busca utilizou as palavras chave: representação do conhecimento, organização da informação/conhecimento, ontologias e teorias classificatórias e catalográficas.

Foram identificados 2.884 pesquisadores citados (sem repetições) para um total de 4.023 referências. Consideraram-se aqueles que apresentaram cinco citações ou mais, em maior número de trabalhos, totalizando 35 pesquisadores citados, partindo-se do pressuposto que, aproximadamente, uma citação por ano, já representaria a evidência de relevância dentro da área.

Com estes dados, foi construída a matriz de ocorrência (115x35), que se refere aos citantes e citados, no software Excel. Gerou-se a matriz quadrada e simétrica de tamanho 35x35 de autores cocitados. Foi usada a função matemática "SOMAR PRODUTO", cujo resultado constituiu a frequência de cocitação entre os autores mais citados nos 115 trabalhos analisados.

Após esse processo, foi utilizado o *software Ucinet* para a construção da rede de cocitação de autores. Para melhor visualização da rede, utilizou-se, no referido *software,* a ferramenta MDS (*Multidimensional Scaling*) como opção de *layout*, na medida em que ela aproxima na disposição visual os autores mais similares em relação às frequências de cocitação. O *layout* tem como finalidade colocar indivíduos na rede, de modo que a distância entre eles reflita sua proximidade em relação aos valores de cocitação (OLIVEIRA, 2013).

Os dados obtidos foram analisados e contextualizados, obedecendo-se ao propósito de delinear a estrutura intelectual da área.

Apresentação e análise dos resultados

A Tabela 1 apresenta a relação dos 35 pesquisadores citados em maior número de trabalhos, com pelo menos cinco citações. Para contabilização dos autores mais influentes foram desconsideradas as autocitações, pois "não representam o impacto de um trabalho (ou autor) sobre outro." (FREITAS, 1997, p.126).

Tabela 1 – Pesquisadores mais citados no período de 2004 a 2013

PESQUISADORES	N° trabalhos citados
Hjørland, B.	**29**
Beghtol, C.	**20**
Olson, H.A.	**17**
Svenonius, E.	**16**
Broughton, V.	**15**
Mai, J.-E.	**15**
Ranganathan, S.R.	**15**
Albrechtsen, H.	**14**
Dahlberg, I.	**13**
Gnoli, C.	**12**
Koch, T.	**12**
Chan, L.M.	**11**
Zeng, M.L.	**11**
Vickery, B.C.	**11**
Bates,M.	**10**
Jacob, E.K.	**10**
Chen, H.	**9**
Vizine-Goetz, D.	**9**
Foskett, D.J.	**8**
Lopez-Huertas, M.J.	**8**
Poli, R.	**8**
Markey, K.	**8**

Golub, K.	**7**
Kipp, M.E.I.	**7**
Campbell, D.G.	**7**
McIlwaine, I.C.	**7**
Smiraglia, R.	**7**
Buckland, M.	**6**
Slavic, A.	**6**
Tennis, J.T.	**6**
Tudhope,R.	**6**
Green,R.	**5**
Hill,L .L.	**5**
Smith, B.	**5**
Spink, A.	**5**

Em relação aos pesquisadores mais citados, destaca-se Hjørland, que apresenta em suas pesquisas uma amplitude temática que atinge a totalidade das linhas teóricas em Ciência da Informação, principalmente, relacionadas à Análise de Domínio, abordagem metodológica utilizada em pesquisas na área.

Destaque também para Beghtol, Olson e Svenonius, com respectivamente 20, 17, 16 pesquisas citadas. Beghtol direciona seus estudos para as questões relativas aos sistemas globais de organização e representação do conhecimento, principalmente sobre acesso, cultura e ética. Olson ressalta a importância crítica da classificação na organização da informação, o uso de uma linguagem ou termo medieval sobre conceitos de gênero, raça e etnia, que pode ser prejudicial no sentido de que marginaliza ou exclui. Já Svenonius é amplamente conhecida por suas pesquisas referentes a teorias sobre controle bibliográfico, especialmente catalogação, classificação e indexação.

A pesquisadora Dahlberg, com 13 citações, apresenta-se como grande teórica da organização do conhecimento, tendo desenvolvido a teoria do conceito na década de 1970. Em seguida Gnoli, o pesquisador mais produtivo no período, aparece com um total de 12 citações, seguido por Koch que concentra em suas pesquisas aspectos da teoria da informação e comunicação digital.

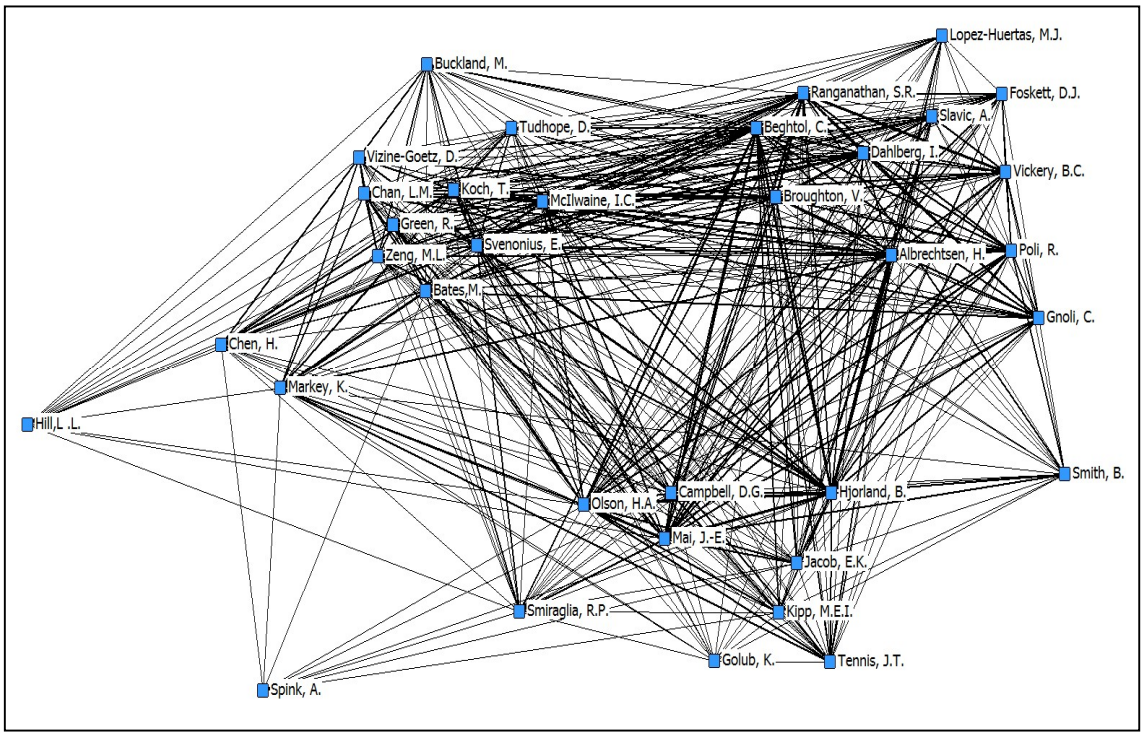
As pesquisadoras Chan e Zeng aparecem com 11 citações cada uma e desenvolveram trabalhos acerca de esquemas de metadados, que podem ser criados ou implementados em bibliotecas digitais. Já o pesquisador Bates, citado em 10 trabalhos, contribuiu com estudos na área de design de sistema de informação centrada no usuário. Muitos dos seus trabalhos têm sido amplamente citados e

utilizados, incluindo artigos sobre seus conceitos e táticas de busca de informação na interface do usuário e do sistema.

Do conjunto de pesquisadores mais citados, a maioria é da área da Ciência da Informação. No entanto, pesquisadores com formação inicial em outras áreas, principalmente da Ciência da Computação, usaram outros referentes para formalizar e fundamentar as teorias acerca da organização do conhecimento e da informação.

A rede de cocitações, apresentada na Figura 1, foi gerada no *software Ucinet*, em que a espessura dos segmentos, representa a frequência de cocitação entre os pesquisadores.

Figura 1. Rede de cocitação entre os pesquisadores em Organização do Conhecimento



Fonte: Elaboração dos autores

Em relação à estrutura geral da rede, destacam-se três grandes grupos, similares em relação às frequências de cocitação e aglutinados pelo MDS. Observa-se que os autores principais mais cocitados são aqueles encontrados no eixo Hjørland, Albrechtsen, Mai, Beghtol e Dahlberg, cujos trabalhos constituem a base das principais teorias em OC, destacando-se a importância dos temas Teoria da Classificação Facetada, Teoria do Conceito, Ontologias entre outros.

Destacam-se os autores centrados no clique entre Hjørland, Mai, Albretchesen e Beghtol, que juntos desenvolveram pesquisas nas temáticas sobre Organização do Conhecimento (OC) e Representação do Conhecimento (RC).

Observa-se a inexistência de pesquisadores brasileiros, no entanto, os pesquisadores que estão contemplados entre os mais citados são a base para as fundamentações teóricas e metodológicas das pesquisas em OC no Brasil.

Apesar de configurar uma rede de um período de estudo recente, observa-se a presença de pesquisadores clássicos, que não só contribuíram para a construção da temática em Organização do Conhecimento, como também para área da Ciência da Informação. Como exemplo, destaca-se o bibliotecário e matemático Shiyali Ramamrita Ranganathan, mais conhecido pela elaboração das Cinco Leis Clássicas da Biblioteconomia (1931), consideradas fundamentais para a base da biblioteconomia moderna. Nesse mesmo limiar, o pesquisador belga Paul Otlet foi considerado pai da documentação por estabelecer sistemas de classificação, como a Classificação Decimal Universal (CDU).

Alguns pesquisadores advindos da área de tecnologias tiveram menor destaque, caso de Spink, Markey e Hill, que também apresentaram em seus estudos abordagens de sistemas e recuperação da informação on-line.

Vizine-Goetz, em atividades de pesquisas atuais, foca a aplicação dos Requisitos Funcionais para Registros Bibliográficos (FRBR), modelos e ferramentas para catalogação automatizada. Também é membro da International Federation of Library Associations - IFLA (Federação Internacional de Associações e Instituições Bibliotecárias) e da ISKO.

Os pesquisadores Olson, Beghtol, Campbell, Lopez Huertas e Sevenonius são estudiosos das questões culturais em OC, enquanto Vickery, Foskett, Slavic, Broughton e Mcllwaine representam o grupo de estudiosos ingleses ligado às teorias de classificação. A díade de pesquisadores Gnoli e Poli desenvolveu estudos de abordagem de ontologias. Atribuem a Dahlberg o papel de fonte estruturante na classificação bibliográfica.

Svenonius, Zeng, Bates e Green representam um grupo americano ligado a teoria de indexação e catalogação de assunto e aos sistemas de classificação da Biblioteca do Congresso (*Library of Congress Classification - LCC)* e CDD.

É possível observar um grupo de estudiosos envolvidos com a questão epistemológica da OC, filosofia da ciência e também de estudos acerca de novos paradigmas no contexto da Ciência da Informação representado por Hjørland, Jacob, Mai, Tennis, Smiraglia e Gnoli. Neste conjunto, Hjørland, Mai, Smiraglia e Tennis têm estudos voltados às metodologias de análise de domínio.

Considerações finais

Esta pesquisa ocupou-se em identificar os autores mais citados e cocitados na produção científica em "Organização do Conhecimento", no periódico K.O., no período de 2004 a 2013, a fim de caracterizar seu domínio.

Do conjunto dos mais citados, a maioria é formada por pesquisadores com pesquisas consistentes em Ciência da Informação. As cocitações presentes e destacadas por segmentos mais fortes na rede, mostram os pesquisadores, as temáticas, a comunidade discursiva e a estrutura social e cognitiva do domínio científico em questão.

No entanto, observa-se a ausência de pesquisadores de renome e destaque na OC entre os mais citados, como Lancaster e Bliss, que em períodos distintos, contribuíram com estudos e pesquisas que garantiram a base para a construção da teoria e prática na Indexação e resumos*: teoria e prática*, a maior contribuição do autor se dá no âmbito da teoria da classificação e da teoria sistêmica.

REFERÊNCIAS

FREITAS, M. H. de A. Oito anos de Transinformação. **Transinformação**, Campinas, v. 9, n. 3, set./dez., 1997.

OLIVEIRA, E. F. T. de. *Análise de Domínio em "Estudos Métricos" no Brasil*: produção, impacto e visibilidade em âmbito nacional e internacional. 2013. 193 f. **Tese** (Livre-Docência em Estudos Métricos em Produção e Organização da Informação). Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2013.

WHITE, H, H.D.; GRIFFITH, B. Author cocitation; a literature measure of intellectual structure. **Journal of the American Society for Information Science & Technolog***y*, v. 32, n. 2, p. 163-171, 1981.

WHITE, H.D.; MCCAIN, K.W. Visualizing a discipline: an author co-citation analysis of Information Science, 1972-1995**. Journal of the American Society for Information Science**, v. 49, n. 4, p. 327-355, 1998.

Abordagens de classificação nos trabalhos apresentados ao GT2 do ENANCIB: classificação profissional e classificação não-profissional

Walter Moreira
Universidade Estadual Paulista
walter.moreira@marilia.unesp.br

Isabela Santana de Moraes
Universidade Estadual Paulista
isabela@marilia.unesp.br

1 Introdução

A teoria da classificação, na condição de princípio orientador e questionador dos processos de organização e representação do conhecimento, tem sido abordada em diversos trabalhos apresentados em eventos científicos da ciência da informação. Conquanto variem as temáticas desses eventos, a problemática referente à classificação, mesmo que com outros nomes, é presença marcante. Isso ocorre por conta do protagonismo da ação de classificação no processo cognitivo, de modo geral, e na organização do conhecimento, de modo especial em bibliotecas.

Desse modo, observando-se a polissemia da expressão “classificação”, essa pesquisa tem como proposta geral a compreensão dos conceitos de “classificação profissional” e de “classificação não-profissional”. Nesse caso, a terminologia também é objeto de investigação. Independentemente, contudo, do consenso terminológico, há duas acepções diferentes de classificação que são efetiva e dialeticamente complementares. O ponto de vista adotado na dicotomia “classificação profissional” e “classificação não-profissional” é o da ciência da informação. Desse modo, a classificação elaborada por um pesquisador em veterinária, por exemplo, para a organização dos conceitos, objetos ou seres em seu campo de estudo, será tomada como uma classificação não-profissional. Parte-se do pressuposto de que nesse caso, esse pesquisador tem como propósito primeiro a descoberta do conhecimento, sem intenções necessárias de generalizações teóricas a respeito da classificação como processo ou como produto. Nesses casos, a teoria da classificação, quando

observada, é utilizada em seu caráter puramente aplicado, isso é, essa classificação não visa ao estudo da teoria pelo próprio avanço da teoria.

Propôs-se como objetivos para este trabalho: a) discutir os conceitos de classificação profissional e não-profissional à luz da teoria da classificação e b) identificar e descrever a presença dessa temática na literatura da ciência da informação brasileira.

Utilizou-se como material os trabalhos apresentados ao Grupo de trabalho "Organização e representação do conhecimento", GT 02, publicados nos anais do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB), com cobertura retrospectiva de cinco anos, isso é, de 2010 a 2014.

O ENANCIB é um evento científico anual promovido pela Associação Nacional de Pesquisa e Pós-graduação em Ciência da Informação (ANCIB) e se configura como um dos espaços mais relevantes e significativos para a discussão e para consolidação da ciência da informação no Brasil. O evento é organizado tematicamente em onze grupos de trabalho.

Utilizou-se como estratégia de busca a expressão "classific*", para que se pudesse localizar o termo "classificação" em suas variações de número e em suas formas adjetivadas, como "classificação bibliográfica", "classificação documentária" etc. A busca foi realizada diretamente no *corpus* documental, isso é, no conjunto de trabalhos apresentados ao GT 2 no período coberto, utilizando-se recursos automáticos de identificação de ocorrências da expressão de busca diretamente nos arquivos. O *corpus* documental foi composto de 187 trabalhos, com a seguinte distribuição anual: 2010 (n=29); 2011 (n=29); 2012 (n=34); 2013 (n=40) e 2014 (n=55).

Para a composição do *corpus* de análise, foram observadas a presença, simultânea ou não, do termo "classificação" nos campos de título e palavras-chave. Esses campos foram escolhidos em função de seu maior poder representacional. Seguindo-se este critério foram encontrados 29 trabalhos.

Após a realização de leitura exploratória, com a finalidade de compreender a temática geral dos trabalhos, analisou-se o conjunto de referências bibliográficas utilizadas com a finalidade de determinar a ocorrência de duas categorias de análise: a) referencial clássico da teoria da classificação/organização e representação do conhecimento (TC/ORC) e b) referencial de "outros" campos ou subcampos.

As seções apresentadas na sequência, visam a relacionar a classificação bibliográfica ao campo maior da organização e representação do conhecimento (seção 2), bem como apresentar os seus fundamentos teórico (seção 3) e conceitual (seção 4). Essa breve fundamentação teórico-conceitual estabelece as bases conceituais para a análise e discussão dos resultados, objeto da seção 5. As considerações finais da pesquisa são apresentadas na seção 6.

2 A classificação no âmbito da organização e representação do conhecimento

O campo da organização e representação do conhecimento (ORC) é estruturado em duas ações (organizar e representar) em torno de um objeto (conhecimento). Essas ações são operacionalizadas com recurso de instrumentos, processos e produtos (FUJITA, 2008). Observando-se ainda o seu caráter teórico-conceitual, a ORC ocupa-se também das bases epistemológicas subjacentes às referidas ações e preocupa-se com o desenvolvimento, a aplicação e a avaliação de sistemas de organização do conhecimento.

Tendo-se em vista a amplitude do objeto (o conhecimento), cabe observar os sentidos complementares que podem ser atribuídos à organização do conhecimento (HJORLAND, 2003; 2008). Num sentido amplo, relaciona-se à organização estrutural e social do conhecimento de modo a atender ao sistema de disciplinas e profissões. Em sentido estrito, compreende atividades como “[...] descrição documentária, indexação e classificação realizadas em bibliotecas, bases de dados bibliográficos, arquivos e outros tipos de ‘instituições de memória’, por bibliotecários, arquivistas, especialistas em informação, especialistas em determinados assuntos [...]” (HJORLAND, 2008, p. 86, tradução livre).

As atividades teórico-práticas de organização e representação do conhecimento, tais como a classificação e a indexação, estão, desse modo, fortemente conectadas como a organização social do conhecimento. Não haveria sentido, aliás, em se manter “instituições de memória” que não dialogassem com o que Dahlberg denominou “conhecimento em ação” (DAHLBERG, 1993; GUIMARÃES, 2008).

3 Teoria da Classificação

O processo de classificação é apontado como recurso necessário e indispensável ao homem em sua relação de conhecimento com o mundo que o cerca

(BROUGHTON, 2004; BATLEY, 2005) e se manifesta quase sempre de modo não totalmente consciente. Para Jacob (2010), não é a classificação, mas o processo cognitivo de categorização que é ubíquo. A distinção ainda não é pacífica e os conceitos de categorização e de classificação possuem realmente muitos traços de similaridade (JACOB, 2004).

Adotando-se a perspectiva de que conhecer é estabelecer relações entre conceitos, requer-se antes, portanto, como tarefa cognitiva, que tais conceitos sejam identificados em suas características e, desse modo, observados no que se refere às suas semelhanças e diferenças. Assim, a identificação do conceito ocorre por meio de processos de classificação que evidenciam o contraste ou a identidade entre os conceitos. A classificação, nessa acepção, refere-se ao agrupamento de coisas que são semelhantes e diz respeito à imposição de algum tipo de estrutura sobre a compreensão da realidade (BATLEY, 2005).

Essa estrutura (referente ao senso comum) rebate-se também num modelo de representação, mesmo que não formalizado e ainda que elaborado *ad hoc*, apenas para a compreensão de um determinado fato ou fenômeno, num nível de compartilhamento conceitual reduzido. Em se tratando de sistemas de classificação bibliográfica, nos quais o caráter representacional é mais evidente e necessário ao seu aspecto pragmático, orientado à construção de representações do conhecimento e de objetos informacionais, a representação requer níveis mais elaborados de formalização consensual que atendam aos interesses do sistema como um todo, incluindo classificadores e usuários.

Nos sistemas de classificação, a noção de agrupamento refere-se à ideia de divisão ou de separação por semelhanças e diferenças, a ordenação, por sua vez, está relacionada ao conceito de sortimento, no sentido de misturar e combinar coisas diferentes. A noção de ordenação diz respeito também à ideia de representação simbólica dos elementos e da sequência que guardam. A representação nesse caso cumpre duas funções: permite que o elemento retirado de sua posição seja recolocado e possibilita a identificação do local “correto” de inserção de um novo elemento, preservando-se a lógica do sistema (BROUGHTON, 2004).

A história da teoria da classificação pode ser pontuada pelo pioneirismo de alguns projetos cujos processos de elaboração atuam como marcos teórico-conceituais, dentre eles a Classificação Decimal de Dewey, publicada em 1876, e a

Classificação Decimal Universal, de Otlet e La Fontaine, em 1905. Esses sistemas, foram atualizados e evoluíram ao longo do tempo (a CDD, por exemplo, está em sua 23ª edição), sendo ainda amplamente utilizados, apesar de algumas limitações que lhes são apontadas.

Nos anos trinta, as contribuições de Bliss (*Bibliographic Classification 1* e, posteriormente, *Bibliographic Classification 2*) e de Ranganathan (*Colon Classification)* modernizaram a teoria da classificação com a abordagem facetada em sistemas analítico-sintéticos. Isso ocorreu pela identificação mais flexível e mais precisa das relações entre as classes, pelo uso de notações mais sintéticas, pelo uso de facetas e pela aplicação consistente da ordem de citação.

A abordagem facetada da classificação foi ampliada, nos anos cinquenta, pelos trabalhos do *Classification Research Group* (CRG) e influenciou diversos outros sistemas de classificação (inclusive os citados CDD e CDU) flexibilizando-os e dotando-os de caráter analítico-sintético, em oposição à característica mais rígida dos sistemas puramente enumerativos.

4 O conceito de classificação

De início, é preciso compreender que toda classificação é arbitrária e deve ser observada em seu caráter de relativização. A expressão “arbitrário” é ambígua em língua portuguesa e possibilita pelo menos três usos, conforme registra o iDicionário Aulete: “1. Que depende do arbítrio ou vontade de quem decide; que não tem regras estabelecidas (medidas arbitrárias); 2. Que segue a sua vontade, sem consideração pela opinião ou necessidade dos outros; DESPÓTICO; 3. Não necessário nem obrigatório; FACULTATIVO” (LACERDA, 2015). O sentido que se quer dar à arbitrariedade nesse contexto está amparado na compreensão que Foucault (2000) tem desse termo, mais próxima das acepções 1 e 2 registradas no dicionário:

> […] não se pode conhecer a ordem das coisas ‘na sua natureza isoladamente’, mas, sim, descobrindo aquela que é a mais simples, em seguida aquela que é a mais próxima para que se possa aceder necessariamente, a partir daí, até as coisas mais complexas (FOUCAULT, 2000, p. 73).
>
> […] uma coisa pode ser absoluta sob certo aspecto e relativa sob outro; a ordem pode ser ao mesmo tempo necessária e natural (em relação ao pensamento) e arbitrária (em relação às coisas), já que uma mesma coisa, segundo a maneira como a consideramos, pode ser colocada num ponto ou noutro da ordem (FOUCAULT, 2000, p. 74)

O caráter de arbitrariedade da classificação, como se pode ver, não a dispensa de apresentar, em qualquer nível e para qualquer finalidade, interconsistência lógica entre todos os seus componentes.

Em busca da compreensão teórica e metodológica a respeito da classificação, Beghtol publicou no periódico *Knowledge Organization,* em 2003, um artigo em que compara alguns aspectos da "classificação profissional" e da "classificação *naïve*". O texto rendeu um acalorado debate no mesmo periódico: Hjorland e Nicolaisen (2003) reagiram ao uso da expressão *naïve*. Beghtol (2004) esclareceu-lhes alguns aspectos e o texto recebeu ainda uma tréplica, escrita por Nicolaisen e Hjorland (2004). Seis anos depois o debate ainda vibrava, com o trabalho de Jacob (2010).

Beghtol (2003; 2004) chama de "profissional" a classificação orientada para a recuperação da informação e de "*naïve*" a classificação realizada para fins de descoberta de conhecimentos. Conforme essa distinção, a classificação para fins de recuperação da informação (profissional) é baseada no consenso e na garantia literária, assentada em concepções mais estáveis e mais largamente compartilhadas. Exatamente por isso suas revisões são mais demoradas. A classificação para descoberta de conhecimento (*naïve*), por sua vez, relaciona-se à descoberta do novo, para isso utiliza como recurso um novo – e mais provisório que o anterior – esquema de classificação, isso é, uma nova hipótese de organização. Os métodos utilizados tanto pela classificação profissional quanto pela classificação *naïve* são similares, conforme Beghtol (2003).

As classificações científicas e acadêmicas, reagiram Hjorland e Nicolaisen (2004), não são ingênuas. Os autores consideram equivocada a compreensão que Beghtol apresentou a respeito das relações entre, de um lado, a classificação acadêmica e científica e, de outro, a classificação feita pela biblioteconomia e pela ciência da informação. A classificação bibliográfica, defendem, é amplamente dependente das classificações acadêmicas e científicas, ainda que não se lhes preste a devida atenção, conforme afirmam: "Nós, na Biblioteconomia e Ciência da Informação, deveríamos obviamente estar preocupados com princípios e métodos de classificação generalizados. Parece, no entanto, muitas vezes, que ignoramos o trabalho feito por cientistas, filósofos e acadêmicos" (HJORLAND; NICOLAISEN, 2004, p. 57, tradução livre).

A temática de Beghtol (2003; 2004), alguns textos que lhe deram origem (DAVIES, 1989; KWASNIK, 1999; RAFFERTY, 2001), bem como as questões levantadas nos textos que debateram sua proposta (HJORLAND; NICOLAISEN, 2004; NICOLAISEN; HJORLAND, 2004; JACOB, 2010, KWASNIK; FLAHERTY, 2010), motivaram a proposta dessa investigação no sentido de explorar neste trabalho (com aprofundamento em posteriores) como o campo da ORC no Brasil tem compreendido as relações entre a "classificação profissional" e a "classificação *não-profissional"*, observando-se, inclusive, a necessidade de investimento em análise conceitual desses termos na área da ciência da informação.

5 Resultados e discussão

A distinção "referencial TC/ORC" e "referencial outros", conforme propõem os objetivos do trabalho, não deve ser tomada como indício de demarcação de exclusividade do campo, reserva de domínio ou qualquer demonstração de xenofobia disciplinar. Essa dicotomia serve apenas para balizar a análise e verificar os rumos que a pesquisa sobre classificação segue no *corpus* de análise. Desse modo, é preciso considerar a subjetividade e os limites desse tipo de categorização.

A categoria "outros" refere-se aos textos que não são identificados classicamente com a TC/ORC, mas que não lhes necessariamente estranhos. Inclui, por exemplo, textos seminais da ciência da informação e que, nessa condição, dialogam com o referencial da TC/ORC. De certo modo, aliás, todos os textos citados pelos autores em seus trabalhos visam a promover o diálogo com a TC/ORC.

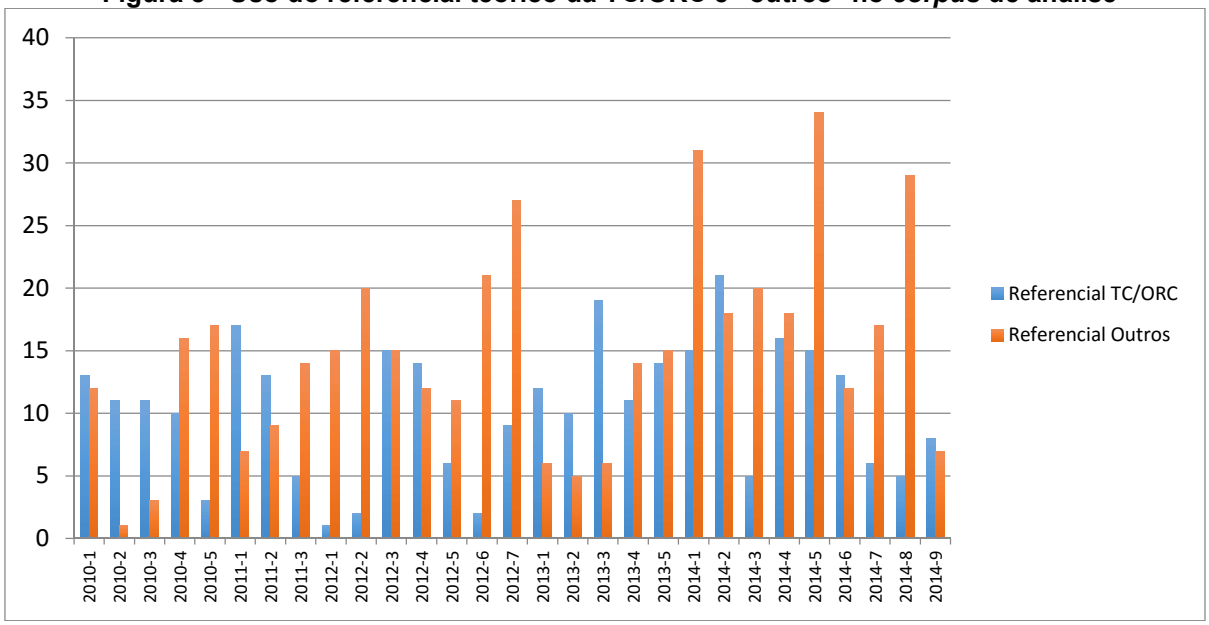
Para o processo de categorização das referências foram observados: o título, a natureza do item (livro, artigo de periódico, trabalhos publicados em congresso etc.) e a afiliação dos autores. Como exemplos, pelo número de ocorrências, podem-se citar: a *Teoria da classificação*, de Requião Piedade, como representativo de TC/ORC e o *Métodos e técnicas de pesquisa social*, de Antonio Carlos Gil, como representativo de "outros".

O número total de artigos analisados foi de 29, com a seguinte distribuição anual: 2010 = 5; 2011 = 3; 2012 = 7; 2013 = 5; 2014 = 9. A Figura 1 apresenta a síntese geral dos dados.

O número médio de referências por artigo é de 25,3, sendo a moda igual a 25. Esse conjunto de referências está distribuído do seguinte modo, em média: 10,4 referências TC/ORC por trabalho e 14,9 referências "outros".

No conjunto geral dos dados, há predominância de referencial “outros”, o que pode revelar alguma tendência ainda a ser verificada em trabalhos posteriores. Considerando-se o número total de referências utilizadas nos 29 trabalhos (n=734), o subtotal de referências TC/ORC representa 41,1% do total (n=302) e o subtotal de referências “outros” equivale a 58,9% do total (n=432).

Figura 8 - Uso do referencial teórico da TC/ORC e "outros" no *corpus* de análise



Fonte: Dados da pesquisa, 2015

Os cinco trabalhos com maior incidência de referencial “outros” abordam temáticas de classificação mais próximas da classificação não-profissional, conforme a conceituação apresentada anteriormente, e referem-se à análise e ao desenvolvimento de propostas de classificação no campo da música, da comunicação e da ciência e tecnologia (Tabela 1).

Tabela 1 - Trabalhos com maior porcentagem de referências "outros"

Trabalho	Foco	Referências TC/ORC	Referências “outros”	% outros	Total referências
2012-1	música	1	16	93,8	17
2012-6	ciência e tecnologia	2	21	91,3	23
2012-2	comunicação	2	20	90,9	22
2014-8	ciência, tecnologia e inovação	5	29	85,3	34
2010-5	comunicação	3	17	85	20

Fonte: Dados da pesquisa, 2015

Tais estudos, quando trazidos para a discussão num evento da área de ciência da informação, já não são mais totalmente identificados com o conceito de classificação profissional em sentido estrito. Observou-se, para fins de ordenação desses trabalhos, além do referencial utilizado, a origem dos problemas que os motivaram.

6 Considerações finais

A abordagem do problema, observando-se os limites desse trabalho, é exploratória e ainda precisa ser ampliada para que se possam produzir interpretações mais precisas. Há, contudo, alguns aspectos quantitativos que parecem confirmar a maior aproximação entre a temática da classificação não-profissional e a utilização de referencial complementar ao referencial mais circunscrito à teoria da classificação.

Observa-se também (Figura 1) que há uma tendência crescente no diálogo entre a teoria da classificação e o referencial bibliográfico de outras áreas ou outros subcampos da ciência da informação. A observação mais acurada desse fenômeno pode auxiliar no esclarecimento terminológico dos conceitos de classificação, profissional ou não, e pode ajudar a esclarecer sobre as diferenças de propósitos e os relacionamentos entre as diferentes modalidades de classificação, ultrapassando-se de modo mais efetivo os limites da teoria da classificação no âmbito da organização e representação do conhecimento.

De modo geral, acredita-se que esse tipo de abordagem contribui para que que sejam evitadas abordagens reducionistas da classificação bibliográfica.

Referências

BATLEY, Sue. **Classification in theory and practice**. Oxford: Chandos, 2005.

BEGHTOL, Clare. Classification for information retrieval and classification for knowledge discovery: relationships between "professional" and "naïve" classifications. **Knowledge organization**, v. 30, n. 2, p. 64-73, 2003.

_____. Response to Hjorland and Nicolaisen. **Knowledge Organization**, v. 31, n. 1, p. 62-63, 2004.

BROUGHTON, Vanda. **Essential classification**. New York: Neal-Schuman, 2004.

DAHLBERG, Ingetraut. Knowledge organization: its scopes and possibilities. **Knowledge organization**, n. 20, p. 211-222, 1993.

DAVIES, Roy. The creation of new knowledge by information retrieval and classification. **Journal of Documentation**, v. 45, n. 4, p. 273-301, 1989.

FOUCAULT, Michel. **As palavras e as coisas**: uma arqueologia das ciências humanas. São Paulo: Martins Fontes, 2000.

FUJITA, M. S. L. Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção científica do ENANCIB no período de 2005 a 2007. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 1, p. 1-32, 2008.

GUIMARÃES, José Augusto C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). **Revista Ibero-americana de Ciência da Informação** (RICI), v.1 n.1, p.77-99, jan./jun. 2008.

HJORLAND, Birger; NICOLAISEN, Jeppe. Scientific and scholarly classifications are not "naïve": a comment to Beghtol(2003). **Knowledge Organization**, v. 31, n. 1, p. 55-63, 2004.
HJORLAND, Birger. Fundamentals of knowledge organization. **Knowledge Organization**, v. 30, n. 2, p. 87-111, 2003.

_____. What is knowledge organization (KO)? **Knowledge Organization**, v. 35, n. 2/3, p.86-101, 2008.

JACOB, Elin K. Classification and categorization: a difference that makes a difference. **Library Trends**, v. 52, n. 3, p. 515-540, winter 2004.

_____. Proposal for a classification of classifications built on Beghtol's distinction between "naïve classification" and "professional classification". **Knowledge Organization**, v. 37, n. 2, p. 111-120, 2010.

KWASNIK, Barbara H. The role of classification in knowledge representation and discovery. **Library Trends**, v. 48, n. 1, p. 22-47, 1999.

_____.; FLAHERTY, Mary Grace. Harmonizing professional and non-professional classifications for enhanced knowledge representation. In: GNOLI, Claudio; MAZZOCCHI, F. (Eds.). **Paradigms and conceptual systems in knowledge organization**: proceedings of the 11th International ISKO Conference, 23-26 feb. 2010, Rome, Italy. Wurzburg: Ergon-Verlag, 2010. p. 229-235.

LACERDA, Carlos Augusto (Ed.) **iDicionário Aulete**. [s.l.]: Lexicon Editora Digital, 2015. Disponível em: <http://www.aulete.com.br/arbitrário>. Acesso em: 27 maio 2015.

NICOLAISEN, Jeppe; HJORLAND, Birger. A rejoinder to Beghtol (2004). **Knowledge organization**, v. 31, n. 3, p. 199-201, 2004.

RAFFERTY, P. The representation of knowledge in library classification schemes. **Knowledge organization**, v.28, n.4, p.180-191, 2001

O tema comportamento informacional no âmbito da ISKO - International Society for Knowledge Organization: estudo referente ao período de 2004 – 2014

Marli Vitor da Silva
Universidade Estadual
Paulistamarli_biblio@hotmail.com

Helen de Castro Silva Casarin
Universidade Estadual Paulista
helen-cs@uol.com.br

1 Introdução

O estudo do comportamento informacional de indivíduos e o uso que os mesmos fazem da informação obtida, é essencial para fundamentar Ciência da Informação, pois várias questões da área, como a organização da informação e a representação dos documentos, são subordinados a ele (ALBRECHTSEN E HÖJRLAND, 1997).

Hjorland (1994) aponta que os indivíduos enquanto usuários da informação são também produtores da informação. A produção da informação é também relevante aos estudos relacionados à área de organização da informação. Deste modo, pode-se considerar os estudos sobre comportamento informacional relevantes para a área de organização da informação enquanto subárea da Ciência da Informação. A partir deste pressuposto, pode-se fazer a seguinte pergunta: como os estudos da área, em particular aqueles que são apresentados nas reuniões da ISKO, têm abordado a temática comportamento informacional?

Assim, este trabalho teve como objetivo verificar como o tema comportamento informacional tem sido discutido e tratado na literatura sobre organização da informação, especificamente no âmbito da International Society for Konwledge Organization. Este evento foi escolhido por ser um dos mais importantes da área de organização da informação e abrigar as principais questões discutidas pela comunidade acadêmica referentes a este tema.

2 A organização da informação e o comportamento informacional

A organização da informação é um processo que diz respeito à descrição física e de conteúdo dos recursos informacionais, tanto tradicionais quanto eletrônicos (BRÄSCHER; CAFÉ, 2008).

Ressalta-se que, de acordo com Garcia Marco (1995), uma teoria abrangente de organização da informação deve ser construída a partir de referenciais de diferentes áreas, incluindo "Estudar as características dos discursos especializados em geral, e os cientistas em particular" (BARITÉ, 2001, p. 57). Deste modo, nota-se que o estudo das características e comportamento de grupos de indivíduos de diferentes áreas do conhecimento que compõem comunidades discursivas são relevantes para a organização da informação.

Assim, é importante que os métodos de organização da informação levem em consideração o comportamento de busca e uso da informação dos usuários em diferentes contextos e domínios do conhecimento (HJORLAND, 1994, p. 91), já que os “instrumentos, conceitos, significados, estruturas de informação, necessidades informacionais e critérios relevantes estão refletidos nas comunidades discursivas” (HJORLAND, 2002, p. 258, TRADUÇÃO NOSSA).

Para Hjorland (1997) o comportamento dos usuários deve ser analisado sob a ótica de uma situação científica em uma determinada área, visto que a investigação de “micro eventos” de comportamento de busca de informação não traz resultados expressivos para a área de organização da informação, pois, não refletem a realidade sobre o comportamento de busca de usuários pertencentes a um determinado grupo (a uma comunidade discursiva). Para o autor, não se pode propiciar aos usuários o acesso a fontes de informação sem que haja algum conhecimento do que é relevante ou não para uma determinada comunidade.

Então, conforme Hjorland (2002, p. 432), pode-se afirmar que são fundamentais à análise de um determinado domínio e sua caracterização os estudos dos usuários da informação, já que permitem identificar as necessidades informacionais de indivíduos em diferentes comunidades, possibilitando, assim, “[...] que um domínio seja organizado segundo preferências, comportamentos ou modelos mentais de seus usuários” (GUIMARÃES, 2015, p. 5).

Assim, Wilson (2000, p. 49, tradução nossa) define comportamento informacional (Information Behavior) como “[...] a totalidade do comportamento humano em relação às fontes e canais de informação, incluindo tanto a busca passiva, quanto a busca ativa, além do uso da informação”.

Conforme Wilson (2000), o termo comportamento informacional engloba outros dois, que se complementam: a) comportamento de busca de informação (Information Seeking Behavior), que diz respeito aos diferentes procedimentos adotados pelos indivíduos para localizar e acessar as informações para sanar as suas necessidades informacionais e b) comportamento de busca em sistemas de informação (Information Search Behavior) está relacionado às interações do usuário com os sistemas de informação baseados em computador.

3 Metodologia

Para atender ao objetivo do trabalho proposto utilizou-se a abordagem de análise de domínio. Tal abordagem se mostra importante como aporte metodológico já que permite caracterizar e avaliar uma determinada ciência, pois permite a identificação das condições pelas quais o conhecimento científico e constrói e se socializa (GUIMARÃES, 2015). Por meio da análise de domínio torna-se possível verificar o que é efetivamente importante ou significativo em um dado campo, de tal modo que aspectos como tendências, padrões, processos, agentes e seus relacionamentos possam ser identificados e analisados (DANUELLO, 2007, APUD GUIMARÃES, 2015).

Desta forma, para a identificação dos estudos de organização da informação que abordam a temática de comportamento informacional foram consultados os anais dos congressos internacionais da ISKO - International Society for Knowledge Organization.

O período contemplado neste estudo compreende os anos de 2004 a 2014, que abrange os congressos realizados em Londres (Inglaterra), Viena (Áustria), Montreal (Canadá), Roma (Itália), Mysore (Índia) e Cracóvia (Polônia). A coleta de dados foi realizada por meio de consulta aos anais impressos (2004 e 2014) e por meio de busca nos anais digitais (2006, 2008, 2010 e 2012). As buscas foram feitas utilizando as palavras-chave: information behavior (our); information seeking behavior (our); information search behavior (our). Primeiramente, foram verificados o título e o

resumo dos trabalhos. Quando o trabalho parecia ser relevante para o estudo, foi verificado o conteúdo do texto na íntegra.

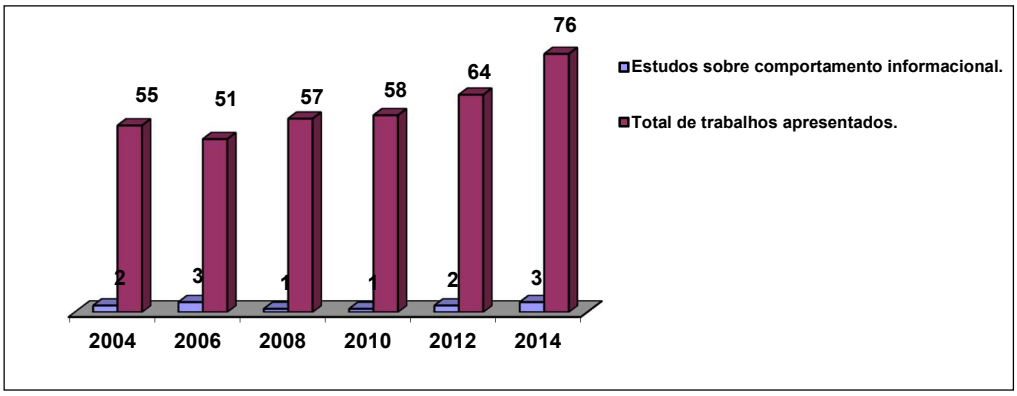
A análise dos trabalhos selecionados foi realizada por meio de Análise de Conteúdo (BARDIN, 2010). Para tanto, foram estabelecidas as seguintes categorias de análise: incidência de trabalhos sobre comportamento informacional; nacionalidade dos autores que têm se dedicado ao tema; metodologia utilizada; âmbito do estudo (comportamento informacional, comportamento de busca de informação e comportamento de busca em sistemas de informação).

4 Apresentação e discussão dos resultados

O total de trabalhos reunidos nos anais das edições da ISKO Internacional no período abrangido neste estudo (2004-2014) é da ordem de 230 trabalhos. Deste total, foram localizados 12 trabalhos que abordam a temática de comportamento informacional, ou seja, 5,2 % do total, revelando que esta é uma temática pouco abordada nos estudos da área.

A incidência de trabalhos sobre a temática ao longo do período estudado, embora pequena, tem sido constante ao longo do período estudado, variando entre dois e três estudos a cada edição do evento, com exceção das edições de 2008 e 2010, nas quais foi localizado apenas um trabalho. Esses dados podem ser visualizados no Gráfico 1.

Gráfico 1: Incidência de trabalhos sobre comportamento informacional nos anais da ISKO Internacional (2004-2014)



Assim, pode-se afirmar que a incidência de trabalhos apresentados sobre o tema de comportamento informacional, em comparação ao total de trabalhos apresentados no evento, parece não ser muito significativo, embora seja constante.

Em relação à origem dos autores dos trabalhos selecionados, verificou-se que estes são oriundos de vários países. Dentre esses, os Estados Unidos foram os que mais se destacaram, pois seis dos 12 trabalhos identificados são de autores estadunidenses (MILLER (2004), PAJARILLO, (2006); LEE, (2006); KIPP, (2008); TILLEY, LA BARRE (2010); TILLEY, LA BARRE, (2012). Os demais países ficaram entre duas apresentações (MATERSKA (2014); KAMINSKA, PULAK (2014) da Polônia) ou somente uma apresentação (SHIRI, REVIE (2004) da Inglaterra), (WILLIAMSON (2006) do Canadá), (BERNAOUI, HASSOUN (2012) da Nigéria); AKERELE, DAVID, OSOFIAN (2014) da França).

Ressalta-se que duas pesquisadoras americanas, Kathryn A. La Barre e Carol L. Tilley, foram as que mais se destacaram, pois apresentaram trabalhos sobre o tema em duas edições do evento, a saber: em Roma (2010) e Mysore (2012).

As metodologias utilizadas para coleta de dados nos trabalhos não variaram muito. Os métodos mais utilizados foram: o experimental em que sujeitos eram convidados a realizarem de tarefas de busca de informação (SHIRI, REVIE (2004); MILLER (2004); WILLIAMSON (2006); LEE (2006); KIPP (2008)), entrevista semi-estruturada (PAJARILLO, (2006); LEE (2006); TILLEY, LABARRE (2010); (TILLEY, LABARRE (2012) e questionário (MILLER (2004); BERNAOUI, HASSOUN (2012); KAMINSK, PULAK (2014).

Destaca-se que quatro dos 12 trabalhos identificados utilizaram a combinação de métodos diferentes de coleta de dados em suas pesquisas, conforme demonstra o Quadro 1.

Quadro 1 – Métodos de coletas de dados utilizados nos trabalhos analisados

TRABALHOS APRESENTADOS	*MÉTODOS DE COLETA DE DADOS UTILIZADOS*
PAJARILLO, Edmund JY (2006)	Entrevista semi-estruturada e Grupo focal
MILLER, Danielle H. (2004)	Questionário e Tarefa de busca
LEE, Hur-Li (2006)	Entrevista semi-estruturada e Tarefa de busca
TILLEY, Carol L.; LA BARRE, Kathryn A. (2010)	Entrevista semi-estruturada e Análise de facetas
TRABALHOS APRESENTADOS	***MÉTODOS DE COLETA DE DADOS UTILIZADOS***
PAJARILLO, Edmund JY (2006)	Entrevista semi-estruturada e Grupo focal
MILLER, Danielle H. (2004)	Questionário e Tarefa de busca
LEE, Hur-Li (2006)	Entrevista semi-estruturada e Tarefa de busca
TILLEY, Carol L.; LA BARRE, Kathryn A. (2010)	Entrevista semi-estruturada e Análise de facetas
TRABALHOS APRESENTADOS	***MÉTODOS DE COLETA DE DADOS UTILIZADOS***
PAJARILLO, Edmund JY (2006)	Entrevista semi-estruturada e Grupo focal
MILLER, Danielle H. (2004)	Questionário e Tarefa de busca
	Entrevista semi-estruturada e Tarefa de busca

LEE, Hur-Li (2006)	
TILLEY, Carol L.; LA BARRE, Kathryn A. (2010)	Entrevista semi-estruturada e Análise de facetas

Com relação à abrangência dos estudos, de acordo com a definição de Wilson (1999) pôde-se constatar que a maioria dos estudos podem ser categorizados como sendo de micro-níveis de comportamento informacional, ou seja, de comportamento de busca e comportamento de busca em sistema de informação, como demonstra o Quadro 2.

Quadro 2: Categorização dos trabalhos de organização da informação nas áreas de Comportamento Informacional

ÁBRANGÊNCIA	***PESQUISAS***
Comportamento informacional (*Information behavior (our)*)	PAJARILLO, Edmund JY. (2006)
Comportamento de busca de informação (*Information seeking behavior (our)*)	TILLEY, Carol L.; LA BARRE, Kathryn A. (2010)
	TILLEY, Carol L. LA BARRE, Kathryn A. (2012)
	BERNAOUI, Radia; HASSOUN, Mohamed. (2012)
	KAMINSKA, Aneta; PULAK, Irena. (2014)
	SHIRI, Ali Asghar; REVIE, Crawford (2004)
	MILLER, Danielle H. (2004)

Comportamento de busca em sistemas de informação (*Information search behavior (our)*)	WILLIAMSON, Nancy J. (2006)
	LEE, Hur-Li (2006)
	KIPP, Margaret E. I. (2008)
	AKERELE, Olubunmi; DAVID, Amos; OSOFISAN, Adenike. (2014)
	MATERSKA, Katarzyna (2014)

5 Considerações finais

O presente estudo revelou que apenas 27,6 % dos trabalhos abordam o tema comportamento informacional, ou seja, tem o foco voltado para o usuário, o que demonstra uma carência de pesquisas dessa temática na área.

Os dados levantados também revelam que, apesar de poucos, há certa regularidade na produção de pesquisas sobre a temática de comportamento informacional, haja vista que no período estudado a temática esteve sempre presente em pelo menos um trabalho.

Os resultados revelaram ainda que as questões mais estudadas são aquelas relativas ao comportamento dos indivíduos quando estão buscando ativamente informações (Information Seeking Behavior), bem quando estão em contato direto com os sistemas de busca de informação (Information Search Behavior). Estes estudos estão voltados tanto para a melhoria dos próprios sistemas, quanto para o aperfeiçoamento das técnicas utilizadas pelos indivíduos na recuperação de informação relevante. A temática de comportamento informacional (Information Behavior), que é o mais abrangente e inclui os comportamento de busca e o comportamento de busca em sistemas de informação, tem sido pouco explorada pela área de organização da informação, demandando mais pesquisas na área.

Referências

AKERELE, O.; DAVID, A.; OSOFISAN, A. Using the concepts of case based reasoning and basic categories for enhancing adaptation to the user's level of knowledge in decision support system. In: **Thirteenth International ISKO Conference - Knowledge organization in the 21st century: between historical patterns and future prospects**, 2014, Kraków, Poland, May 19-22. Proceedings of the Thirteenth International ISKO Conference. Würzburg: Ergon, 2014.

ALBRECHTSEN, H.; HJÖRLAND, B. Information seeking and knowledge organization. **Knowledge Organization**, v. 24, n. 3, p.136-144, 1997.

BRÄSCHER, M.; CAFÉ, L. **Organização da Informação ou Organização do Conhecimento**, 2008. Disponível em: <http://www.ancib.org.br/media/dissertacao/1835.pdf>. Acesso em: 08 abr. 2014.

BARDIN, L. **Análise de conteúdo**. Lisboa: Edições 70, 2010. 281 p.

BARITÉ, M. Organización Del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (Org.). **Educação, universidade e pesquisa**. Marília: Unesp – Marília – Publicações, 2001. p. 35-60.

BERNAOUI, R.; HASSOUN, M. User Expectations, Reality and Delineation of Agricultural Information Systems in the Maghreb. In: Twelfth International ISKO Conference - Categories, contexts and relations in knowledge organization, 2012, Mysoure, India, August 6-9, 2012. **Proceedings of the Twelfth International ISKO Conference.** Würzburg: Ergon, 2012.

GARCIA MARCO, J. Los contenidos y la secuencia docente de la Organización y Representación del Conocimiento: una propuesta interdisciplinar. **Organización del Conocimiento en Sit. Inf. Y Docum.,** v. 1, p. 219-228, 1995.

GUIMARÃES, J. A. C. **A análise de domínio com perspectiva metodológica em organização da informação**, 2015, 12 p.

HJORLAND, B. Nine Principles of knowledge organization. **Advances in knowledge organization**, v. 4, p. 91-100, 1994.

HJORLAND, B. **Information Seeking and subject representation: an activity-theoretical approach to Information Science**. London: Greenwood Press, 1997. 214 p.

HJORLAND, B. Epistemology and the socio cognitive perspective in information science. **Journal of the American Society for Information Science and Technology**, v. 53, n. 4, p. 257-270, 2002.

KAMINSKA, A.; PULAK, I. Knowledge organization in a digital learning environment in the experriences of pedagogy students. In: In: Thirteenth International ISKO Conference - Knowledge organization in the 21st century: between historical patterns and future prospects, 2014, Kraków, Poland, May 19-22. **Proceedings of the Thirteenth International ISKO Conference.** Würzburg: Ergon, 2014.

KIPP, M. E. I. Searching with Tags: Do Tags Help Users Find Things? In: tenth International ISKO Conference - Culture and identity in knowledge organization, 2008, Montréal, Canada, August 5-8. **Proceedings of the Tenth International ISKO Conference.** Würzburg: Ergon, 2008.

LEE, Hur-Li Navigating Hierarchies vs. Searching by Keyword: Two Cultural Perspectives. In: Ninth International ISKO Conference - Knowledge organization for a global learning societ, 2006, Vienna, Austria, July 4-7. **Proceedings of the Ninth International ISKO Conference**. Würzburg: Ergon, 2006.

MATERSKA, K. Faceted navigation in search and discovery tools. In: In: Thirteenth International ISKO Conference - Knowledge organization in the 21st century: between historical patterns and future prospects, 2014, Kraków, Poland, May 19-22. **Proceedings of the Thirteenth International ISKO Conference**. Würzburg: Ergon, 2014.

MILLER, D. H. User perception and online online catalogue: public library OPAC users “think aloud”. In: Eighth International ISKO Conference - Knowledge organization and the global information society, 2004, Londres, England, July 13-16. **Proceedings of the Eighth International ISKO Conference**. Würzburg: Ergon, 2004.

PAJARILLO, E. JY. A qualitative research on the use of knowledge organization in nursing information behavior. In: Ninth International ISKO Conference - Knowledge organization for a global learning societ, 2006, Vienna, Austria, July 4-7. **Proceedings of the Ninth International ISKO Conference.** Würzburg: Ergon, 2006.

SHIRI, A. A.; REVIE, C.End-user interaction with thesauri: an evaluation of cognitive overlap in search term selecton. In: Eighth International ISKO Conference - Knowledge organization and the global information society, 2004, Londres, England, July 13-16. **Proceedings of the Eighth International ISKO Conference**. Würzburg: Ergon, 2004.

TILLEY, C. L.; LA BARRE, K. A. New models from old tools Leveraging an understanding of information tasks and subject domain to support enhanced discovery and access to folktales. In: Eleventh International ISKO Conference - Paradigms and conceptual systems in knowledge organization, 2010, Rome, Italy, February 23-26. **Proceedings of the Eleventh International ISKO Conference**. Würzburg: Ergon, 2010.

TILLEY, C. L.; LA BARRE, K. A. What if they build it and no one comes? Balancing Full-Text Access and User Tasks. In: Twelfth International ISKO Conference - Categories, contexts and relations in knowledge organization, 2012, Mysoure, India,

August 6-9. **Proceedings of the Twelfth International ISKO Conference**. Würzburg: Ergon, 2012.

WILLIAMSON, N. J. Knowledge structures and the Internet: Progress and Prospects. In: Ninth International ISKO Conference - Knowledge organization for a global learning societ, 2006, Vienna, Austria, July 4-7. **Proceedings of the Ninth International ISKO Conference.** Würzburg: Ergon, 2006.

WILSON, T. D. Human Information Behavior. **Information Science Research**, v. 3, n. 2, 2000.

A representação de assunto no escopo da Arquivologia: uma análise de artigos científicos nacionais e internacionais

Luciane Paula Vital
Universidade Federal de Santa Catarina
luciane.vital@ufsc.br

Graziela Martins de Medeiros
Universidade Federal de Santa Catarina
graziela.m@ufsc.br

Leolibia Luana Linden
Universidade Federal de Santa Catarina
leolibialuana@gmail.com

Marisa Brascher
Universidade Federal de Santa Catarina
marisa.brascher@ufsc.br

1 Introdução

O termo informação arquivística é recente na literatura e deriva de uma mudança de paradigma, em que o tratamento passa do documento para a informação representada nele. Essa mudança aproxima, ainda mais, a Arquivologia da Ciência da Informação, quando se propõe a estudar os fenômenos relacionados à informação, e faz com que, consequentemente, questões relacionadas ao acesso, como o tratamento descritivo e temático recebam destaque.

Nesse tratamento, vale ressaltar a representação de assunto, que carece de aprofundamento teórico no escopo da Arquivologia, considerando os estudos desenvolvidos na Ciência da Informação, sem desconsiderar as especificidades do documento arquivístico, como o fato de manterem relações entre si, que necessitam

ser explicitadas por meio de princípios arquivísticos. É fundamental a consideração dos princípios que regem a Arquivologia para não incorrer na descaracterização de seus documentos.

Assim, esse estudo buscou verificar como a representação de assunto em documentos de arquivo é tratada nas publicações da área de Ciência da Informação. Para isso, realizou-se buscas em bases de dados nacionais e internacionais, verificando-se os diferentes aspectos apresentados pelos autores. Esse trabalho se constitui em uma primeira aproximação exploratória da temática, sem a pretensão de realizar um aprofundamento da discussão sobre as abordagens teóricas identificadas.

2 Organização e representação de documentos arquivísticos

A organização e representação da informação na arquivística são desenvolvidas, especificamente, nos processos de classificação e descrição. A classificação diz respeito à ordenação física e intelectual dos documentos, de forma hierarquizada e com explicitação das relações entre eles. Já a descrição arquivística diz respeito à organização e representação da informação objetivando gerir e recuperar os documentos. A definição de descrição arquivística apresentada na *General International Standard Archival Description (ISAD (G)),* permite afirmar que se constitui em um processo de organização e representação da informação:

> A elaboração de uma acurada representação de uma unidade de descrição e suas partes componentes, caso existam, por meio da extração, análise, organização e registro de informação que sirva para identificar, gerir, localizar e explicar documentos de arquivo e o contexto e o sistema de arquivo que os produziu. (CONSELHO INTERNACIONAL DE ARQUIVOS, 2000, p.4).

No Brasil, a descrição é respaldada por uma norma internacional, já citada, a ISAD (G), e uma norma nacional, a Norma Brasileira de Descrição Arquivística (NOBRADE), publicada em 2006. Em relação a norma internacional, a Nobrade acrescenta a área de pontos de acesso e indexação de assuntos, em relação a norma internacional e sob esse aspecto nos parece que avança na consolidação do entendimento da descrição arquivística como atividade de Organização e Representação da Informação (ORI). Porém, de acordo com Oliveira (2012, p.64) “É evidente que a área dialoga com a questão dos pontos de acesso de forma pouco

consistente quando os elementos extrapolam o admitido mais tradicionalmente [...]". Os elementos tradicionais apontados pela autora são, por exemplo, produtor, data e série. Nesse caso, a autora deixa de mencionar a autoria, os dados de publicação e o próprio título, o que vai de encontro ao que a norma considera - que os itens documentaispodem conter a autoria. Observa-se nessa questão uma imprecisão terminológica.

Ribeiro (1996, p.114), ao se referir à realidade portuguesa, e após pesquisa em instituições arquivísticas, conclui que há necessidade de pontos de acesso por assunto, e afirma que: "carecemos, também, quer a nível nacional, quer internacionalmente, de princípios orientadores em matéria de indexação, que sejam facilmente aplicáveis à realidade dos Arquivos."

Nessa perspectiva, esse trabalho apresenta um levantamento dos trabalhos que tem tratado da representação de assunto em documentos de arquivo, com o objetivo de contribuir com a discussão sobre essa temática..

3 Metodologia

A pesquisa é caracteriza como qualitativa e exploratória, considerando que a análise de dados visa identificar artigos que tratam de aspectos conceituais e aplicados da representação de assunto em documentos arquivísticos.

Realizamos buscas avançadas nas base de dados LISA, BRAPCI e BENANCIB, nos campos título, resumo e palavras-chave. Utilizamos os seguintes termos: representação de assunto, indexação, análise de assunto e representação da informação, combinados com o termo arquivologia. Não foi realizado recorte temporal.

No total foram identificados e analisados 17 artigos que tratam da temática, sendo 10 na LISA, 4 na BRAPCI e 3 no BENANCIB. Ressalta-se que foram desconsiderados os documentos que não obtivemos acesso ao texto completo. Os artigos analisados foram classificados em duas categorias:

Categoria 1 - Artigos conceituais: Documentos que apresentam aspectos conceituais referentes à representação de assunto (e/ou análise de assunto e indexação) em documentos arquivísticos e/ou no contexto dos arquivos.

Categoria 2 - Artigos com aplicação prática e pesquisas em andamento: Documentos que apresentam aplicações metodológicas e pesquisas em andamento referentes ao processo de representação de assunto em documentos arquivísticos.

O quadro 1 apresenta as referências, em ordem alfabética, identificadas e analisadas nesse trabalho.

AGUIAR, Francisco Lopes de; KOBASHI, Nair Yumiko. Organização e representação do conhecimento: perspectivas de interlocução interdisciplinar entre ciência da informação e arquivologia. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...** Florianópolis: UFSC, 2013.

CAPONE, Vera Lucia Punzi Barcelos; CORDEIRO, Rosa Inês de Novais. O arquivo fotográfico ilustrativo dos trabalhos geográficos de campo do Instituto Brasileiro de Geografia e Estatística (IBGE): proposta de uma matriz para análise documentária da paisagem. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 14., 2013, Florianópolis. **Anais...**Florianópolis: UFSC, 2013.

CZECK, R L H. Archival MARC records and finding aids in the context of end-user subject access to archival collections. **American Archivist**, v.61, p. 426-440, out., 1998.

DOOLEY, J. M. Subject indexing in context. **American Archivist**, Estados Unidos, v. 55, n. 2, p. 344-354, Apr. 1992.

ESTEBAN NAVARRO, M. A.. La representación y la organización del conocimiento en los archivos. In: ACTAS DEL I ENCUENTRO DE ISKO-ESPAÑA. **Anais...** Madrid, 4 y 5 de noviembre de 1993. Zaragoza: Universidad, Librería General. 1995. p. 65-90.

GAGNON-ARGUIN, L. Documentary analysis 2: Current problems relating to documentary analysis. **Archives (Quebec),** Quebec, v. 28, n. 3, p. 23-41, 1997.

DELGADO GOMEZ, Alejandro. Metadatos y descripción como representación del contexto archivístico. **Ibersid:** Revista de sistemas de informacion y documentacion, Zaragoza, Espanha, p. 157-164,2007.

HUDON, M. Indexing and documentary languages in archival environments in the era of new information technologies. **Archives (Quebec),** Quebec, v. 29, n. 1, p. 75-98, 1998.

LARADE, S P; PELLETIER, J M. Mediating in a neutral environment: gender-inclusive or neutral language in archival description. **Archivaria,** v.35,abr., p. 99-109, 1993.

LEVESQUE, Michel L'indexation: luxeounecessite? **Archives (Quebec)**, Quebec, v. 33, n.1, p. 17-45, 2001.

LOPEZ, André Porto Ancona; CARVALHO, Pedro Davi Silva. A classificação arquivística por assunto em documentos fotográficos: o exemplo do arquivo público do Distrito Federal.**Perspectivas em Gestão & Conhecimento**, v. 3, n. 2, p. 271-279, 2013

SANTOS, Maria José Veloso da Costa. Correspondência científica de Bertha Lutz: um estudo de aplicação da lei de Zipf e ponto de transição de Goffman em um arquivo pessoal. **Ponto de Acesso,** v. 3, n. 3, p. 317-326, 2009.

SCHENKOLEWSKI-KROLL, Silvia; TRACTINSKY, Assaf. Archival description, information retrieval, and the construction of thesauri in Israeli archives. **Archival Science,** v. 6, n.1, p. 69-107, mar. 2006,.

SIMÕES, Maria da Graça; FREITAS, M. Cristina V. de. A classificação em arquivos e em bibliotecas à luz da teoria da classificação: pontos de convergência e de divergência. **Ponto de Acesso**, Salvador, v. 7, n. 1, p. 81-115, 2013.

SILVA, Márcio Bezerra da; SOUSA, Dulce Elizabeth Lima de; BANDEIRA, Pablo Matias. A representação temática em documentos arquivísticos: o caso da indexação documental realizada pelos alunos de arquivologia da UFPB. **InCID: Revista de Ciência da Informação e Documentação**, São Paulo, v. 3, n. 2, 2012.

SOUSA, Renato Tarciso Barbosa de. A representação da informação: classificação e indexação automática de documentos de arquivo. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 15., 2014, Belo Horizonte. **Anais...** Belo Horizonte: UFMG, 2014.

YAKEL, Elizabeth. Archival representation. **Archival Science,** v.3, n.1, p. 1-25, 2003.

Quadro 1: Referências analisadas

4 Análise e discussão dos resultados

Dentre os 17 artigos analisados, 11 foram classificados na categoria 1, sobre os quais discorremos a seguir.

Dentre os 17 artigos selecionados, 11 foram considerados conceituais. A classificação e a descrição arquivísticas são apresentadas como processos de organização e representação da informação e do conhecimento por Simões e Freitas (2013), Esteban Navarro (1995) e também por Yakel (2003).

Por outro lado, Yakel (2003) define o termo 'representação arquivística' como o mais assertivo, na atualidade, para nomear o trabalho do arquivista de (re)ordenar, interpretar, criar substitutos e 'estruturas' para sistemas de representação.

No processo de organização e representação de documentos arquivísticos, Lopez e Carvalho (2013) ressaltam a necessidade de respeito aos princípios dessa área, especialmente proveniência e organicidade. Nesse sentido, Capone e Cordeiro (2013), por exemplo, apresentam uma proposta de análise de fotografias da paisagem rural estabelecendo categorias que incluem informações de contexto. Essa acepção é compartilhada nos artigos que fizeram parte da análise, tanto entre autores nacionais como internacionais, destacando-se: Gagnon-Arguin (1997), Hudon (1998), Lopez e Carvalho (2013), Capone e Cordeiro (2013), Sousa (2014). Considerar os princípios arquivísticos, para os referidos autores, é garantir a reconstituição histórica e, assim, o valor de prova da documentação.

Outros autores acrescentam a essa discussão o uso das linguagens documentárias. Schenkolewski-kroll e Tractinsky (2006) discutem a necessidade de uma padronização na utilização de linguagens documentárias, especialmente o tesauro, no processo de indexação de documentos arquivísticos. Salientam a necessidade de se ter uma estrutura mínima, adaptada a realidade de cada arquivo, a exemplo das normas de descrição arquivísticas. Os autores apresentam dois tesauros utilizados em arquivos israelenses, no contexto da recuperação da informação em sistemas automatizados, isto é, centrados na questão tecnológica. Gagnon-Arguin (1997) afirma que o uso do vocabulário controlado e tesauro tem sido discutido entre os arquivistas do Quebec e ressalta ser esse um desafio para a área, assim como Dooley (1992), que verifica a necessidade de acesso por assunto em

arquivos de forma mais consistente, considerando também as características dos documentos arquivísticos (forma do material, tempo, lugares geográficos, funções).

Com outro enfoque, a língua também foi apontada por Larade e Pelletier (1993) como um elemento fundamental no processo de representação de documentos em bibliotecas e arquivos. Destacam que a linguagem é parcial, fruto de posicionamentos políticos e interfere no acesso e recuperação da informação. Questiona a forma com que as instituições arquivísticas se posicionam no processo de parcialidade da língua na representação da informação.

Foram recuperados três artigos publicados no Canadá: Gagnon-Arguin (1997), Hudon (1998) e Lévesque (2001), que são convergentes quanto à visão de representação de assunto apresentada, uma vez que a relacionam ao processo de descrição e consideram o assunto como uma forma adicional de acesso.

Gagnon-Arguin (1997) frisa que o tratamento de assunto em arquivo tem sido uma preocupação crescente entre os arquivistas quebequences desde 1986, expressa por meio de publicações sobre o tema e por iniciativas de utilização da indexação de assunto ou de cabeçalhos de assuntos relacionadas às normas de descrição arquivística, como forma de melhorar o acesso aos documentos de arquivo. Nessa publicação observamos a forte preocupação com os usuários e suas formas de busca no processo de recuperação da informação.

Hudon (1998) relaciona, de maneira direta, o processo de indexação de assuntos com a descrição arquivística e os níveis de descrição, informando que qualquer política de indexação será necessariamente ligada à definição de políticas de descrição vigentes. Acrescenta que a organização e estruturação das informação por conteúdo é apenas um acesso adicional. Indica como desafios o estabelecimento de políticas de indexação, sistemas de indexação e linguagens documentárias eficazes pelos arquivistas.

Para Lévesque (2001), os arquivistas concentram-se na descrição, minimizando a importância da indexação. Ou seja, os arquivistas costumam privilegiar outras formas de acesso aos documentos de arquivo que não o seu conteúdo. Ao mostrar exercícios de classificação e descrição, seguidos de discussão teórica, evidencia que o contexto é o ponto de vista mais privilegiado. Porém, na visão do autor, o arquivista deve ajudar o pesquisador ou usuário a obter diferentes formas de acesso aos documentos.

Nos parece que os autores canadenses, especificamente quebequences, são pioneiros quanto ao tema, visto que apontam a existência de publicações e de iniciativas institucionais no país desde 1986. São convergentes quando apontam o acesso ao documento pela proveniência como o principal, no âmbito da arquivística, e apontam uma lacuna quanto aos estudos de representação de assuntos.

No que diz respeito à categoria 2, relacionada às aplicações práticas e/ou pesquisas em andamento, classificamos um total de 6 artigos, sobre os quais passamos a tratar.

Santos (2009) apresenta uma aplicação prática das leis de Zipf e Ponto de Transição de Goffman no processo de representação automática de conteúdos em documentos arquivísticos, no contexto de um arquivo pessoal. Porém, o autor não discute a pertinência do processo de representação de assunto em arquivos e as especificidades desse ambiente.

Também tratando da indexação automática, Silva, Sousa e Bandeira (2012) apresentam uma aplicação prática de indexação e criação de índices com documentos arquivísticos, a partir das orientações da área 8 da Nobrade. Destacam que a representação temática da informação na Arquivologia encontra-se em processo de desenvolvimento técnico e prático. Nesse viés de aplicação de tecnologias, Czeck (1998) analisa o conteúdo de instrumentos de pesquisa (guias, índices, inventários), de acervos arquivísticos, objetivando determinar como o registro MARC pode influenciar na representação dessas informações.

Na categoria pesquisas em andamento, os artigos identificados, publicados no Brasil, foram os de Aguiar e Kobashi (2013) e Sousa (2014). Na Espanha, Delgado Gomez (2007) trata a representação da informação como uma forma de manter, ao longo do ciclo de vida documental, seu valor de prova. Analisa e discute modelos de metadados, como por exemplo a ISO 23081 - Metadados para a gestão de documentos e o *Encoded Archival Description* (EAD), padrão para representação de instrumentos de pesquisa em ambientes *Web,* propondo uma futura aplicação multidimensional.

Aguiar e Kobashi (2013) destacam a possibilidade de interlocução entre a organização e representação da informação, da área de Ciência da informação, e a Arquivologia, em uma área que pode ser denominada de "Organização e Representação da Informação Arquivística (ORIA). O trabalho objetiva sintetizar as

principais contribuições teórico-conceituais e metodológicas advindas da Organização e Representação do Conhecimento como possibilidade nos processos e sistemas de Organização e Representação Arquivística

Já Sousa (2014), apresenta uma pesquisa que tem como objetivo a construção de um modelo para classificação e indexação automática de documentos de arquivo, preservando os princípios de respeito aos fundos e proveniência. Discute o que seria assunto em documentos arquivísticos e faz apontamentos no sentido de que outros aspectos do material de arquivo podem fornecer pistas importantes sobre o assunto do documento, fortalecendo a relação da proveniência com o conteúdo sem ênfase em um ou no outro. Essas pesquisas nos mostram que a problemática está sendo abordada na atualidade, na busca de elementos que contribuam para metodologias de representação de documentos arquivísticos.

6 Considerações finais

Como evidenciam os resultados, a organização e representação da informação arquivística vem sendo discutida com mais ênfase a partir da década de 90. O Canadá tem destaque nos trabalhos teóricos, que discutem a representação de assunto relacionada com os princípios arquivísticos.

As buscas realizadas nas bases de dados que fizeram parte da pesquisa permitiram verificar que o número de documentos que efetivamente, se referem à temática proposta são bastante reduzidos. Verficamos que os artigos selecionados são, em sua maioria, exploratórios, em oposição às discussões conceituais mais aprofundadas. Esse pode ser considerado um dado relevante para a área, visto que denota uma necessidade expressa de realização de pesquisas de cunho teórico-conceitual.

Quanto aos documentos que tratam de aspectos conceituais, percebemos uma preocupação com os aspectos relacionados à relação entre a representação de assuntos e as funções arquivísticas de classificação e descrição. Observamos, ainda, a defesa enfática da consideração dos princípios de organicidade e proveniência em qualquer processo de organização e representação de documentos arquivísticos.

Os trabalhos com aplicações práticas e pesquisas em andamento nos mostram uma tendência ao destaque da tecnologia, como representação automática de conteúdo e padrões de metadados.

Assim, pode-se concluir que o tema é pouco explorado nos artigos recuperados nas bases de dados da área de Ciência da Informação, apesar dos autores ressaltarem a necessidade de aprofundar as discussões e, inclusive, proporem estudos nesse sentido.

Referências

BORKO, H. Information science: what is it**? American Documentation**, v. 19, n. 1, p. 3-5, jan. 1968.

BRASIL. Conselho Nacional de Arquivos. **NOBRADE**: Norma Brasileira de Descrição Arquivística. Rio de Janeiro: Arquivo Nacional, 2006. Disponível em: http://www.conarq.arquivonacional.gov.br/Media/publicacoes/nobrade.pdf. Acesso em: 04 jul. 2012.

CONSELHO INTERNACIONAL DE ARQUIVOS. **ISAD(G):** norma internacional de descrição arquivística. 2.ed. Rio de Janeiro: Arquivo Nacional, 2000. (Publicações técnicas, n. 49). Disponível em: http://www.conarq.arquivonacional.gov.br/Media/publicacoes/isad_g_2001.pdf. Acesso em: 04 jul. 2012.

HEREDIA HERRERA. A. **Archivistica General:** teoría y práctica. 5.ed.Sevilla : Diputación de Sevilla, 1991.

RIBEIRO, F. **Indexação e controle de autoridades em arquivo**. Porto: Câmara Municipal do Porto, Arquivo Histórico, 1996.

_____. Organizar e representar a informação: apenas um meio para viabilizar o acesso? **Ciência e Técnicas do Patrimônio:** Revista da Faculdade de Letras, Porto, v. 4, p. 1-22, 2005.

SILVA, I.O.S. **A organização e a representação do conhecimento no domínio da arquivística.** 2012, 193f. Tese (Doutorado em Ciência da Informação) - Universidade Estadual Paulista, Marília, 2012.

OLIVEIRA, A. L.V. **Descrição e pesquisa:** reflexões em torno dos arquivos pessoais. Rio de Janeiro: Móbile Editorial, 2012.

Análise da metodologia em estudos bibliométricos: uma proposição de indicadores de contexto

Maria Guiomar da Cunha Frota
Universidade Federal de Minas Gerais
guiomar@eci.ufmg.br

Ana Cláudia Ribeiro
Universidade Federal de Minas Gerais
anacribeiro29@gmail.com

1 Introdução

A produção científica é responsável pela consolidação de novos conhecimentos, resultado do processo e desenvolvimento, que ocorrem por intermédio das pesquisas. A ela cabe importante papel no desenvolvimento científico e tecnológico dos países.

A Bibliometria permite analisar e avaliar os processos relacionados à atividade científica. Assim, os estudos bibliométricos vêm se consagrando ao longo dos anos como importante método de análise e, uma das técnicas mais adequadas de estudos para visualização de produção científica. De acordo com Guedes e Borschiver (2005), a Bibliometria é um “conjunto de leis e princípios empíricos que contribuem para estabelecer os fundamentos teóricos da Ciência da Informação”.

A análise da produção do conhecimento, permite criar métodos confiáveis para mostrar à sociedade, como uma determinada área vem se desenvolvendo, suas tendências e o comportamento de uma comunidade. Para Macias-Chapula (1998), a Bibliometria vem se apresentando como uma ferramenta capaz de auxiliar na observação do “estado da ciência e da tecnologia através da produção da literatura científica como um todo, em um determinado nível de especialização”.

A criação de indicadores para as publicações permite à comunidade científica, e aos próprios pesquisadores, refletir sobre a evolução da área e o caminho que estão tomando no que for relativo a determinado assunto ou disciplina. Isso auxilia na

atividade de pesquisa em si, pois, a visualização da área permite o acompanhamento acerca do desdobramento de estudo, cooperando com o desenvolvimento científico.

Pretende-se, de modo mais amplo, investigar a metodologia utilizada nos artigos referentes ao livro HAYASHI, Maria Cristina P. Innocentini; MUGNAINI, Rogério.; HAYASHI, Carlos Roberto Massao (*Orgs*); Encontro Brasileiro de Bibliometria e Cientometria. **Bibliometria e cientometria**: metodologias e aplicações. São Carlos: Pedro & João, 2013. E testar a metodologia na Sur Revista Internacional de Direitos Humanos.

Especificamente as etapas metodológicas para observação dos dados foram os seguintes:

- Recuperar os artigos científicos do livro para utilização nas sucessivas etapas do trabalho.
- Identificar os indicadores mais utilizados nas pesquisas bibliométricas recentes conforme a literatura
- Analisar esses indicadores tendo como balisadores as características relevantes de cada um.
- Refletir sobre a pertinência de criar novos indicadores capazes de apurar a inserção sócio-política de determinada publicação.

A Sur Revista Internacional de Direitos Humanos foi criada em 2002 com o objetivo de aproximar acadêmicos do hemisfério sul que atuam no campo dos direitos humanos e de promover a cooperação destes com agências da ONU. A Revista ainda enquadra segundo a última classificação da Capes em A2 e de acordo com a base de dados a única de Qualis A, com foco em Direitos humanos.

Os países da América do Sul vêm se desenvolvendo consideravelmente em níveis sociais e econômicos nos últimos anos, por isso a importância de discutir, refletir e ampliar o papel dos direitos humanos como elemento central das relações humanas, devendo ser contemplado pelos meios governamentais nas políticas públicas e nas universidades através de debates, pesquisas, publicações e cursos de formação para profissionais de diversas áreas.

3 Pressupostos teóricos

3.1 Direitos Humanos

O Artigo 19º da Declaração Universal de Direitos Humanos de 1946 destaca: “Todo ser humano tem direito à liberdade de opinião e expressão; este direito inclui a liberdade de, sem interferência, ter opiniões e de procurar, receber e transmitir informações e ideiais por quaisquer meios e independentemente de fronteiras” (ONU, 2009). A liberdade de informação é reconhecida como um direito fundamental e pedra basilar de todas as liberdades. O acesso à informação é um direito individual correlato ao direito de liberdade de expressão e autonomia.

A Ciência da Informação tem um papel imprescindível no que tange o Artigo 19º da Declaração Universal dos Direitos Humanos. Configurando como um campo do conhecimento interdisciplinar e socialmente responsável, por meio de sua atuação ou pela produção científica criada para atender as necessidades de informação da sociedade, portanto, uma ciência social que se inter-relaciona com outras áreas do conhecimento, preocupada em corresponder às expectativas e necessidades de informação da sociedade.

Considerando a amplitude e importância dos Direitos Humanos, em termos de diversidade de literatura científica, é razoável pensar que o tema esteja se tornando cada vez mais complexo e heterogêneo. Esses fatores indicam a relevância social e cultural que o campo vem conquistando ao longo dos anos.

Neste sentido, e sob o olhar da Ciência da Informação, o conhecimento precisa ser organizado, de modo que o avanço produzido pelos pesquisadores seja transformado em informação acessível para a comunidade científica (Macias-Chapula, 1998). A atividade científica possui um caráter cumulativo, baseado no consenso. Esse conhecimento acumulado representa a memória coletiva da ciência, e se materializa na forma de literatura técnica-científica. Quando analisadas por meio de técnicas de mensuração da informação podem apontar tendências de desenvolvimento da sociedade, das disciplinas científicas, bem como das áreas de produção e consumo (WORMELL, 1998).

Apesar do fato de pesquisas sobre Direitos Humanos terem se intensificado após da Segunda Guerra Mundial, verifica-se que ainda faltam estudos sobre a produção científica nesta área.

Assim, é oportuno a realização de um diagnóstico dos indicadores de metodologia bibliométrica aplicáveis à essa área e de novos indicadores que possam potencializar a análise do contexto sócio-político. A verificação da pertinência de se criar novos indicadores será realizada a partir de uma análise bibliométrica da Revista Sur de Direitos Humanos.

Ressalta-se que estudos dessa natureza são de fundamental importância tanto para o campo dos Direitos Humanos, como para a Ciência da Informação, pois, afirmam o caráter social e interdisciplinar, bem como de sua vocação para os estudos das propriedades, do comportamento e das forças que regem o fluxo da informação científica.

3.2 Sur Revista Internacional De Direitos Humanos

A Sur Revista Internacional de Direitos Humanos foi criada em 2002 com o objetivo de aproximar acadêmicos do hemisfério sul que atuam no campo dos direitos humanos e de promover a cooperação destes com agências da ONU.

A Revista vêm fortalecendo vínculos entre acadêmicos e defensores do Sul Global dedicados a direitos humanos. A Revista Sur tem como objetivo ampliar as vozes e a participação desses atores dentro das organizações internacionais e universidades. O principal foco do periódico são os países do cone sul e suas normativas e políticas relativas aos Direitos Humanos.

3.3 Bibliometria

Segundo Price (1976) estudar a produção científica traz a possibilidade de conhecer e caracterizar as tendências e lacunas da ciência. A importância da avaliação científica deu-se a partir da invenção da imprensa no século XV, quando houve grande aumento na disponibilidade de textos impressos na Europa, o que desencadeou o crescimento da produção de livros. Esse crescimento de fontes de informação causou impacto na difusão das informações, bem como na avaliação do que era divulgado. Passando aumentar de forma exponencial as publicações, entretanto faltava um sistema de avaliação.

Segundo Macias-Chapula (1998), a Bibliometria permite o estudo dos aspectos quantitativos da produção, da disseminação e do uso da informação registrada e desenvolve padrões e modelos matemáticos para medir esses processos, usando os resultados para elaborar provisões e apoiar tomadas de decisão.

Para realização do estudo bibliométrico é necessário saber quais parâmetros se relacionam melhor com a pesquisa. Tais parâmetros são conhecidos como “leis bibliométricas”. Essas leis são definidas por Vanti (2002, p.153) conforme o quadro a seguir:

Lei de Lotka ou Lei do Quadrado Inverso	**Lei de Zipf ou Lei do Mínimo Esforço**	**Lei de Bradford ou Lei de Dispersão**
Aponta para a mediação da produtividade dos autores, mediante um modelo de distribuição do tamanho da freqüência dos diversos autores em um conjunto de documentos.	Consiste em medir a frequência do aparecimento das palavras em vários textos, gerando uma lista ordenada de termos de uma determinada disciplina ou assunto.	Mediante a medição da produtividade das revistas, permite estabelecer o núcleo e as áreas de dispersão sobre um determinado assunto em um mesmo conjunto de revistas.

Quadro 1 – Leis bibliométricas

Fonte: Vanti (2002, p.153).

Assim, a Bibliometria vem sendo utilizada nas diversas áreas do conhecimento como metodologia para a desenvolvimento de indicadores na avaliação da produção científica. As metodologias associadas aos *softwares* também indicam os caminhos a serem seguidos pela comunidade científica e contribuem com o intercâmbio das informações, bem como na transferência e no gerenciamento do conhecimento em determinado campo de estudo selecionado.

4 Procedimentos Metodológicos

A metodologia adotada é de natureza exploratória e descritiva em duas etapas: 1ª.) abordagem bibliométrica para o levantamento dos principais indicadores incluídos no livro: HAYASHI, Maria Cristina P. Innocentini; MUGNAINI, Rogério.; HAYASHI, Carlos Roberto Massao (*Orgs*); Encontro Brasileiro de Bibliometria e Cientometria. **Bibliometria e cientometria:** metodologias e aplicações. São Carlos: Pedro & João, 2013. e 2ª.) a aplicação da bibliometria aos artigos da Sur Revista de Direitos Humanos (2011-2013).

Os dados obtidos foram sistematizados com o objetivo de analisar primeiro os indicadores dos artigos em bibliometria, visando analisar estes indicadores. Foram no total analisados 13 artigos publicados no livro Bibliometria e Cienciometria: metodologias e aplicações, a posteriori foram analisados 50 artigos da Sur Revista Internacional de Direitos Humanos (2011-2013).

4.1 Análise dos indicadores das pesquisas bibliométricas.

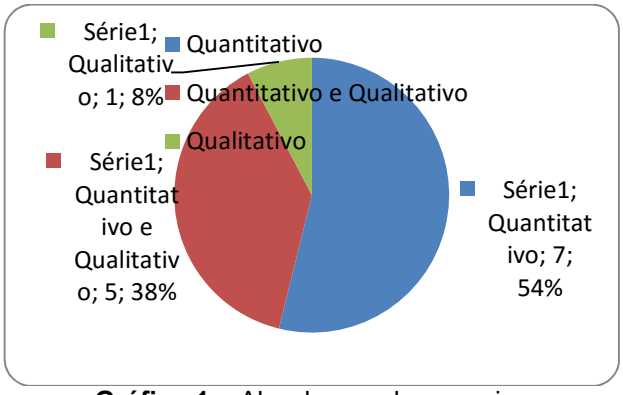


Gráfico 1 – Abordagem da pesquisa

Fonte: Elaborado pelo autor.

O gráfico 1 representa a maior concentração dos artigos com abordagem quantitativa 54%, seguida da abordagem qualitativa 38% e quantitativo e qualitativo 8%.

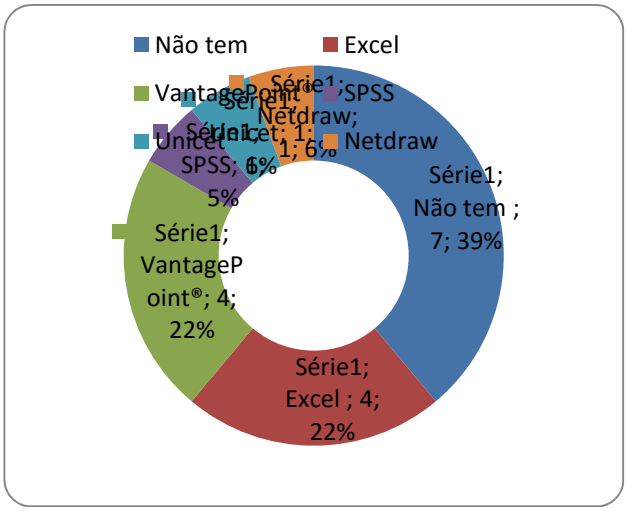


Gráfico 2 – Software
Fonte: Elaborado pelo autor.

No gráfico 2 a maioria não utilizou *software* resultando em 39%; o Excel resultou em 22%, entretanto, não é considerado um *software* bibliométrico; o *software* bibliométrico mais utilizado foi o *VantagePoint*® com 22%; seguidos do *Unicet* e *Netdraw* cada um com 6% e *SPSS* com 5%. Cabe destacar que e o *software* é de suma importância para tratamento dos dados e composição e da pesquisa.

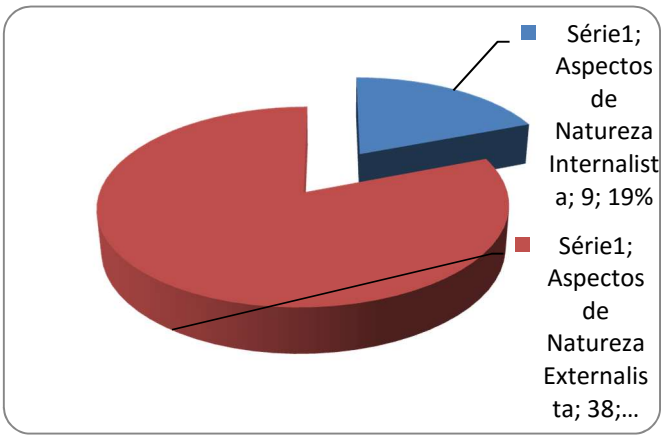


Gráfico 3 – Natureza dos Indicadores Bibliométricos
Fonte: Elaborado pelo autor.

Os aspectos externalistas visam quantificar dados, como:

1. a distribuição temporal e geográfica da produção científica (quando e onde os estudos foram realizados);
2. o comportamento dos pesquisadores e a participação dos centros de pesquisa (quem foram os autores e avaliadores, estudos e em quais instituições foram realizados);

3. as áreas do conhecimento representadas (áreas às quais os estudos estão vinculados).

Quanto aos aspectos internalistas visam caracterizar o conteúdo dos trabalhos científicos, quantificando dados como os termos utilizados pelos pesquisadores para descrever as pesquisas no que se refere as áreas do conhecimento as quais essas pesquisas estão relacionadas e as palavras-chave que indicam os temas abordados pelas pesquisas.

4.2 Titulação dos autores da Sur Revista Internacional de Direitos Humanos (2011-2013).

Indicadores analisados: Graduação, Mestrado e Doutorado dos autores. A partir dessa perspectiva da análise de Titulação dos autores é possível construir um indicador do contexto sócio-político. Como a produção em Direitos Humanos é heterogênea e interdisciplinar, será possível compor as áreas de conhecimento que mais vêm contribuindo com a temática de Direitos Humanos.

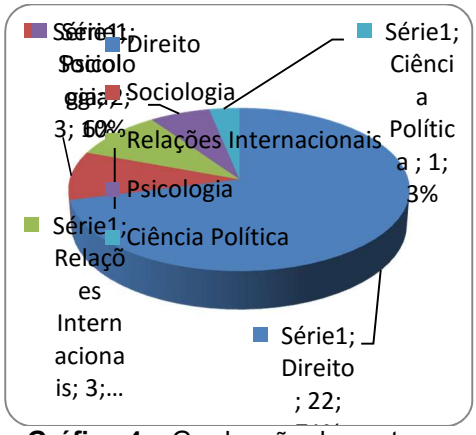


Gráfico 4 – Graduação dos autores

Fonte: Elaborado pelo autor.

Dos 50 artigos analisados 71% dos autores possui a graduação em Direito, seguidos das formações em Relações Internacionais e Sociologia cada um 10%, Psicologia 6% e Ciência Política 3%.

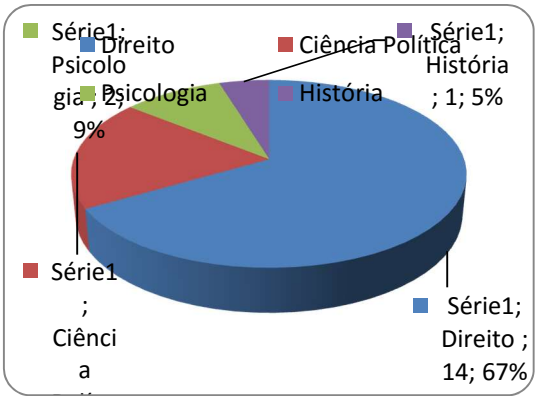


Gráfico 5 – Mestrado dos autores

Fonte: Elaborado pelo autor.

Na formação em Pós-graduação dos 50 artigos analisados, 67% dos autores possui mestrado em Direito, seguidos das formações em Ciência Política 19%; Psicologia 9% e História 5%.

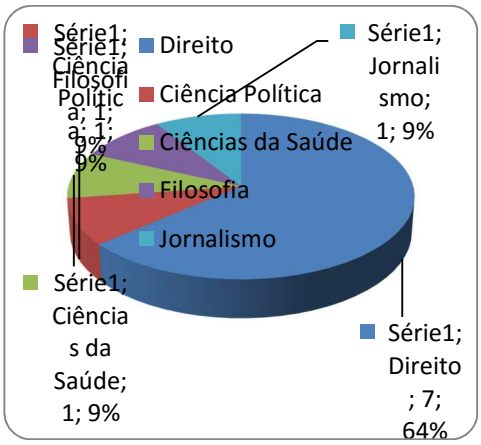


Gráfico 5 –Doutorado dos autores

Fonte: Elaborado pelo autor.

Na formação de Doutoramento dos autores, foram analisados 50 artigos dentre eles, 64% em Direito, seguidos de Ciência Política, Ciências da Saúde e Filosofia cada um com 9%. Estes indicadores são relevantes, pois a formação dos autores demonstra as áreas de conhecimento que vêm apresentando maiores contribuições para a área de Direitos Humanos.

Esta pesquisa trata, não apenas de quantificar e constatar, mas de atribuir sentido aos dados, qualificando-os para que possam ter melhor uso na produção científica em Direitos Humanos e, que a pesquisa reverbere para a sociedade, seja por cada especialidade ou grupo de pesquisa, ou em contextos mais abrangentes,

regionais, nacionais ou mundiais. Segundo Santos e Kobashi (2009) “[...] os estudos métricos se aproximam, cada vez mais, das ciências ditas ‘moles’ (CHS) porque estas últimas oferecem teorias e modelos que permitem interpretar os dados em contextos culturais, políticos, ideológicos e econômicos distintos.”

A criação de novos indicadores para as publicações que permita à comunidade científica, e aos próprios pesquisadores, refletir sobre a evolução da área e o caminho que estão tomando no que é relativo a determinado assunto ou disciplina. Isso auxilia na pesquisa em si, pois, a visualização da área permite o acompanhamento acerca do desdobramento de estudo, cooperando com o desenvolvimento científico.

5 Considerações Finais

A discussão sobre a questão dos indicadores bibliométricos se inicia com a própria missão da bibliometria que é de visualização de determinada área de conhecimento. Faz parte da bibliometria acompanhar o desdobramento, apontar tendências e refletir sobre o próprio objeto de estudo. É necessário estudar as características gerais e específicas de cada área do conhecimento para se estabelecer formas de visualização que contemplem as particularidades de cada área da produção científica.

Há preocupação em testar e estabeler novos indicadores, em virtude da pesquisa que se encontra em fase de desenvolvimento do mestrado junto ao grupo de pesquisa: Memória, repressão e reparação às violações de direitos humanos nos países do cone sul (ECI/UFMG). E assim testar novos indicadores efetivos e que possam contemplar nas análises bibliométricas. Pois, observa-se uma lacuna em pesquisas acerca da produção científica em Direitos Humanos que vão além de uma análise quantitativa.

Deste modo, este trabalho buscou propor um debate de maneira a estimular a investigação inovadora utilizando indicadores bibliométricos que se conformem com aspectos específicos da temática dos Direitos Humanos, preservando a análise do contexto o qual circunda a referida temática.

Referências

BUFREM, L.; PRATES, Y. O saber científico registrado e as práticas de mensuração da informação. **Ciência da Informação***.* Brasília, v. 34, n.2, p.9-25, maio/ago. 2005. Disponível em: < http://revista.ibict.br/ciinf/index.php/ciinf/article/viewArticle/682>. Acesso em: 08 abril de 2014.

GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2010.

GUEDES, V.; BORSCHIVER, S. **Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica**. In: CINFORM – ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6., 2005, Salvador. Anais... Salvador: ICI/UFBA, 2005.

HAYASHI, Maria Cristina P. Innocentini; MUGNAINI, Rogério.; HAYASHI, Carlos Roberto Massao (*Orgs*); Encontro Brasileiro de Bibliometria e Cientometria. **Bibliometria e cientometria:** metodologias e aplicações. São Carlos: Pedro & João, 2013.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 134-140, maio/ago. 1998. Disponível em: < http://revista.ibict.br/ciinf/index.php/ciinf/article/view/342>. Acesso em: 03 março de 2014.

OLIVEIRA, M. de. Origens e Evolução da Ciência da Informação. In: OLIVEIRA, M. de. (org). **Ciência da Informação e Biblioteconomia: novos \conteúdos e espaços de atuação**. Belo Horizonte: Editora UFMG, 2005. p.9-28.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). **Direitos humanos**. Disponível em:<http://www.un.org/en/rights/>. Acesso em: 23 março de 2015.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). **A Declaração Universal dos Direitos Humanos**. 2009. Disponível em:<http://www.dudh.org.br/definicao/documentos/>. Acesso em: 23 março de 2015.

PRICE, D. de S. **O desenvolvimento da ciência**: análise histórica, filosófica, sociológico e econômica. Rio de Janeiro: Livros Técnicos e Científicos, 1976.

SUR Revista internacional de direitos humanos. São Paulo: Conectas, 2014. Disponível em: < http://www.surjournal.org/sobresur19.php>. Acesso em: 10 maio de 2015.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, Brasília, v.31, n.2, p. 152-162, maio./ago. 2002. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-19652002000200016&lang=pt>. Acesso em: 15 maio de 2015.

WORMELL, I. Informetria: explorando bases de dados como instrumentos de análise. **Ciência da Informação**, v. 27, n. 2, p. 210-216. 1998. Disponível em: < http://revista.ibict.br/ciinf/index.php/ciinf/article/viewArticle/367>. Acesso em: 25 março de 2015.

Organização e representação do conhecimento: contribuições aos estudos métricos

Cibele Araújo Marques dos Santos
Universidade de São Paulo
cibeleac@usp.br

1 Introdução

Este estudo tem por objetivo identificar a contribuição da Organização e Representação do Conhecimento (ORC) para os estudos métricos da produção científica em Ciência da Informação no Brasil que posteriormente possibilitem desenvolver metodologias para indicadores temáticos de pesquisa.

A produção científica cresce em proporção ilimitada na sociedade atual devido aos recursos tecnológicos e a multiplicação de ferramentas para divulgação destas informações, incluindo o acesso a informação permitido pela web. Do mesmo modo, percebem-se desafios expressivos quanto ao desenvolvimento de políticas em pesquisa, principalmente nas universidades que são grandes produtoras de informação científica e tecnológica no Brasil.

Neste contexto, a análise da produção científica torna-se recurso fundamental para o planejamento de políticas científicas nacionais e institucionais e os indicadores bibliométricos podem identificar padrões e tendências das áreas. São estas métricas que permitem medidas de atividade, impacto, colaboração e dinâmica da comunicação de resultados de pesquisa que são representados por indicadores, analisando de forma quantitativa os dados de publicação e citação (SANTIN, BRAMBILLA, STUMPF, 2013).

Por outro lado, a análise da temática da produção científica permite uma visão mais qualitativa na medida que com o mapeamento dos temas podemos verificar a cobertura da pesquisa e as lacunas existentes. Porém, as abordagens para este tipo de análise geram debate em relação aos métodos mais adequados para representar um domínio disciplinar a partir de uma estruturação temática, no que se refere aos níveis de subjetividade, validade e atualização do modelo de classificação utilizado (LIBERATORE e HERRERO-SOLANA, 2013). A dispersão decorrente da variabilidade e granularidade de assuntos abordados em um conjunto de artigos de periódicos é grande quando analisamos as palavras-chaves de cada artigo, além de ocorrerem inconsistências, uso de sinônimos, variações linguísticas, bem como critérios diferenciados de nivelamento do tratamento da informação pelos autores, principalmente em áreas interdisciplinares. As ferramentas de área de Organização e Representação do Conhecimento, podem auxiliar na superação destas dificuldades para realizar mapeamentos mais consistentes.

2 Metodologia

Foi realizada pesquisa exploratória em artigos de revistas brasileiras para identificar a metodologia utilizada na definição das temáticas em Ciências da Informação nos estudos métricos.

O levantamento bibliográfico foi realizado através de pesquisa na base Brapci com estratégia de busca, realizada em maio de 2015, que utilizou o campo de palavras-chaves pesquisando os termos: estudos métricos (resultou em 9 artigos), bibliometria (120 artigos), análise bibliométrica (18 artigos), análise temática (2 artigos, apenas um em bibliometria) e mapeamento (15 artigos), sendo que estes resultados trouxeram muitos artigos, nem sempre pertinentes à bibliometria temática.

Os artigos foram selecionados a partir do título e do resumo, visando análise dos recursos da área de Organização e Representação do Conhecimento utilizados pelos autores para realizar os mapeamentos temáticos.

2.1 Estudos métricos: características e temáticas

Existem diversos artigos, trabalhos publicados em anais de eventos, dissertações e teses com abordagens bibliométricas quantitativas e qualitativas que analisam a produção científica em Ciência da Informação a partir de indicadores relacionados aos autores, como afiliação e co-autoria, referências e citações de autores, revistas e co-citações, porém não são muitos os trabalhos que incluem análise temática.

A produção científica, como apontado por Liberatore et al (2007), considerando docentes e pesquisadores em nível de pós-graduação, apresenta-se analisada por fontes, metodologias, temáticas, linhas de pesquisa e formas de difusão. Os autores em trabalho de revisão consistente informam que as teses defendidas nos programas foram estudadas também nas questões relacionadas a seu desenvolvimento, problemáticas, objetos de estudo e modos de citação.

Nesse sentido, Macedo (1987) publicou artigo sobre pesquisa em Ciência da Informação e Biblioteconomia que abordava entre outras questões a temática de teses, sendo que os dados foram coletados nos Catálogos de Dissertação e Teses em Ciência da Informação e Biblioteconomia publicados entre 1982 a 1985 pelo IBICT. Os assuntos identificados para as teses e dissertações foram retirados da indexação realizada pela fonte bibliográfica destes dados, os catálogos citados, e utilizavam tesauro construído a partir do documento “Percepção estrutural da Ciência da Informação e áreas correlatas” com as seguintes divisões: Generalidades; Ensino, pesquisa e atividade profissional; Organização das atividades de informação e de bibliotecas; Estudos da literatura e do documento; Entrada, tratamento, armazenamento da informação; Recuperação e disseminação da informação, Transferência e uso da informação e áreas correlatas. No artigo citado, as teses foram distribuídas por grandes assuntos com titulação e dados geográficos e também analisadas pelos assuntos mais específicos do tesauro.

Esta divisão de assuntos também era utilizada segundo o Instituto Brasileiro de Informação em Ciência e Tecnologia (1984) pela Bibliografia Brasileira de Ciência da Informação que referencia artigos de periódicos na década de 80 do século 20.

Poblacion e Noronha realizaram diversos estudos sobre a produção da pós-graduação no Brasil a partir do Núcleo de Produção Científica da ECA-USP. Em artigo publicado em 2002 analisam a produção caracterizando-a tematicamente pelas linhas de pesquisas institucionalizadas de seus autores nos programas de pós-graduação de

cinco universidades brasileiras e incluíam 22 linhas de pesquisas diferentes que continham temáticas comuns, porém dispersavam os assuntos dos documentos.

Em outro artigo Queiroz e Noronha (2004), mapearam tematicamente as dissertações e teses do programa de pós-graduação da ECA-USP utilizando para classificação temática o *Broad Subject Headings*, a lista de cabeçalho de assunto do *Library and Information Science Abstracts* (LISA), que em sua versão impressa apresenta 19 categorias temáticas subdivididas em assuntos específicos. Informa que cada um dos 114 itens foi classificado em apenas umas das 19 categorias da lista, mais uma subcategoria dentro deste assunto principal, além de uma das linhas de pesquisas do programa. As categorias utilizadas foram: Biblioteconomia e Ciência da Informação; Profissões; Bibliotecas e centros de informação; Uso de bibliotecas e usuários; Materiais; Organização; Prédios de bibliotecas; Tecnologia em bibliotecas; Serviços técnicos; Difusão da informação; Controle bibliográfico; Registro bibliográfico; Armazenamento e recuperação da informação; Tecnologia da informação e comunicação; Leitura; Mídia; Conhecimento e aprendizado; Gerenciamento de registros; Outros assuntos correlatos.

Mueller e Pecegueiro (2001), analisaram artigos da revista Ciência da Informação de 1990-1999, que consideraram como elementos significativos da produção científica da área, indicando os temas predominantes e comparando a frequência destes e os descritores registrados na base LISA, que apresentaram diferenças expressivas. Foi utilizada uma tabela com 9 macrotemas: Ciência da Informação, Biblioteconomia e Documentação; Ensino, atividade profissional e pesquisa; Organização e gerência de atividades de informação, de bibliotecas e centros de pesquisa; e Estudo de usuários, transferência e uso da informação e da biblioteca; Estudos da literatura e do documento; Prédios de bibliotecas; Serviços técnicos; Entrada, tratamento, armazenamento, recuperação e disseminação da informação, Outros assuntos correlatos ou adicionais.

Percegueiro (2002) publicou outro trabalho sobre a temática dos artigos de seis revistas brasileiras em Ciência da Informação que foram classificados pela tabela anterior agora com o acréscimo do tema “Política de informação e política cientifica e tecnológica”. A classificação foi comparada novamente com a ocorrência de temas na base LISA e foram também encontradas diferenças. Neste trabalho, a autora chama a atenção para a oscilação de determinados temas ao longo dos anos como por

exemplo o pico de publicação sobre “Ensino, atividade profissional e pesquisa em 1990 caindo para 5 artigos em 1991 e 2 em 1992. O tema Prédios não teve publicações no período e Serviços Técnicos teve poucos artigos publicados.

3 Organização e Representação do Conhecimento e métricas

A contribuição da área de ORC aos estudos métricos, nas pesquisas destacadas acima, mostra procedimentos de classificação por macrotemas ou por assuntos específicos, o uso do tesauro do IBICT, da lista de termos da base LISA e de vocabulário controlado adaptado para estudos métricos temáticos. A utilização destas linguagens documentárias ocorreu através da indexação e às vezes reindexação dos dados coletados, porém não foram discutidas, na maioria dos estudos, as questões linguísticas e terminológicas envolvidas na escolha dos vocabulários e as metodologias dos procedimentos de análise documentária.

Em estudo sobre a revista Ciência da Informação, Silva (2002), discute rapidamente aspectos relacionados ao vocabulário controlado e questões de linguagem nos índices para análise temática de produção científica. Vanz e Stumpf (2010), discutem os processos de avaliação da produção incluindo fontes de coleta de dados, bases de dados nacionais e internacionais, Google Acadêmico e outros, softwares utilizados, procedimentos para limpeza e padronização de dados, considerados imprescindíveis para eliminar as inconsistências, bem como a organização da produção em áreas do conhecimento. As autoras ponderam que a inserção da produção em grandes áreas torna-se uma forma de evitar a sobreposição de assuntos, como o uso de esquemas de classificação como as 15 divisões do conhecimento do ISI ou as tabelas de áreas do CNPq-CAPES.

Na pesquisa realizada para a aplicação bibliométrica na análise de dissertações e teses, Kobashi e Santos (2010), propõem abordagem interdisciplinar dos Estudos Sociais da Ciência, da Organização e Representação do Conhecimento e dos métodos bibliométricos avançados. Os autores lembram que os estudos bibliométricos fundamentam-se tradicionalmente na descrição e quantificação e propõem combinar abordagem de estudos quantitativos e qualitativos. Nestes estudos, os indicadores de relação são construídos pela co-ocorrência, seja autoria, citações ou palavras, e podem ser aplicados na elaboração de mapas estruturados de

conhecimento e redes de relacionamento, considerando que estes mapas são representações da produção científica da área analisada. Existe também a possibilidade de utilizar conceitos produzidos pela própria área, que traz “garantia literária na geração das representações cartográficas da institucionalização cognitiva de um dado campo do conhecimento”.

A utilização das palavras-chave dos autores nas análises temáticas foi recorrente, sendo que Araujo e Melo (2011) na análise dos artigos publicados na revista Perspectivas em Ciência da Informação, optou por utiliza-las e observou que a falta de controle terminológico das palavras-chave, gera uma “variação imensa na maneira de expressar determinados conceitos da área”. Comentam que, quando foi possível, realizaram a padronização terminológica e que a dispersão dos termos e as diferentes categorias usadas como palavras-chave, impediram uma análise mais aprofundada possibilitando apenas a visualização de um quadro sugestivo de temas abordados.

Por sua vez, Liberatore e Herrero-Solana (2013) analisaram os temas de investigação científica em Ciência da Informação entre 2000 e 2009 estudando artigos de 4 revistas da área, e em sua abordagem metodológica optaram por não adotar um esquema específico, utilizando as palavras-chaves dos autores por entenderem ser uma representação ajustada a percepção dos próprios autores ao seu campo de atuação semântico. Obtiveram uma lista de 2498 palavras pertencentes a 965 artigos, analisaram este universo pelo cálculo da co-ocorrência, fixando um mínimo de 8 ocorrências como uma representação significativa.

Realizaram também nesta pesquisa uma análise dos artigos pelos macrotemas dos grupos de trabalho da ANCIB. A matriz de dados resultante dessas análises em conjunto, foi apresentada por meio de uma rede social cujos nós indicados pelo número de relações entre as ocorrências permitiu identificar o grau e a centralidade destas relações. Nos resultados, os autores chamam a atenção para os componentes da subjetividade dos autores nas temáticas, a questão do nível de consenso na terminologia empregada pela comunidade científica e a dispersão das expressões utilizadas.

4 Considerações finais

Foi realizado um levantamento bibliográfico não exaustivo selecionando artigos que tenham feito mapeamento temático em estudos bibliométricos, analisando na metodologia destes trabalhos os procedimentos da Organização e Representação e Organização do Conhecimento empregados. Observou-se que os autores, normalmente, não se referem ao processo de indexação de forma detalhada com a indicação da metodologia e critérios, por exemplo, mostrando apenas a lista de termos utilizada, se a mesma pertence a um tesauro ou vocabulário institucional, não discutindo também as categorias presentes.

Não houve uma lista de assuntos predominante, mas foram localizados artigos que mostraram questões quanto a linguagem e uso de grandes áreas temáticas nos procedimentos, a importância de uma abordagem interdisciplinar e o uso dos recursos da ORC.

No entanto, os artigos não apresentaram discussões sobre a influência das terminologias utilizadas nos resultados da pesquisa, embora em um trabalho tenha ocorrido uma padronização terminológica e em outro um artigo que trabalhou com as palavras-chaves dos autores, tenha sido ressaltado o aspecto da subjetividade que a escolha envolveu.

Podemos observar que contribuição da ORC para os estudos métricos não se esgota nos pontos levantados e nem nos trabalhos identificados e que existe ainda um campo de pesquisa conjunta que pode ser mais explorado permitindo dessa forma o aprimoramento das metodologias e o aprofundamento dos mapeamentos temáticos em Ciência da Informação.

Referências

ARAÚJO, C. A. A.; MELO, M. O. T. Análise dos quinze anos do periódico Perspectivas em Ciência da Informação. **Perspectivas em Ciência da Informação**, v.16, n.4, p.243-256, 2011.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA. **Bibliografia Brasileira de Ciência da Informação** v.6 (1980;83). Brasília, IBICT, 1984.

KOBASHI, N. Y.; SANTOS, R. N. M. dos. Arqueologia do trabalho imaterial: uma aplicação bibliométrica a análise de dissertações e teses. **Encontros Bibli**, n. esp. p. 106-115, 2008.

LIBERATORE, G.; HERRERO-SOLANA, V. Caracterizacion temática de la investigacion em Ciencia de la Informacion en Brasil en el período 2000-2009. **Transinformação**, v.25, n.3, p.225-235, 2013.

LIBERATORE, G.; HERRERO-SOLANA, V.; GUIMARÃES, J. A. C. Análise bibliométrica do periódico brasileiro Ciência da Informação durante o período 2000-2004. **Brazilian Journal of Information Science**, v.1, n.2, p. 3-21, 2007.

MACEDO, N. D. de. Pesquisa em Ciência da Informação e Biblioteconomia: implicações na pós-graduação; análise temática. **Ciência da Informação**, v.16, n.2, p.129-144, 1987.

MUELLER, S. P. M.; PECEGUEIRO, C. M. P. de. O periódico Ciência da Informação na década de 90: um retrato da área refletido em seus artigos. **Ciência da Informação,** v. 30, n. 2, p.47-63, 2001.

PECEGUEIRO, C. M. P. de. Temática dos artigos de periódicos brasileiros na área da ciência da informação na década de 90. **Transinformação**, v.14, n.2, p.117-131, 2002.

POBLACION, D. A.; NORONHA, D. P. Produção das literaturas “brancas” e “cinzenta” pelos docentes-doutores dos programas de pós-graduação em ciência da informação no Brasil. **Ciência da Informação**, v.31, n.2, p.98-106, 2002.

QUEIROZ, F. M.; NORONHA, D. P. Temática das dissertações e teses em Ciência da Informação no Programa de pós-graduação em Ciências da Comunicação da USP. **Ciência da Informação**, v.33, n.2, p. 132-142, 2004.

SILVA, F. M. e. Análise da revista Ciência da Informação disponibilizada na SciELO a partir do seu vocabulário controlado. **Transinformação**, v.14, n.2, p.133-138, 2002.

VANZ, S. A. de S.; STUMPF, I. R. C. Procedimentos e ferramentas aplicados aos estudos bibliométricos. **Informação & Sociedade**, v.20, n.2, p.67-75.

Apoio: **PROJETO FAPESP 2013-04109-7**

Relativizando os índices h e g: estudo aplicado ao domínio dos Estudos Métricos

Deise Deolindo Silva
Universidade Estadual Paulista
deisedeolindo@hotmail.com

Maria Cláudia Cabrini Grácio
Universidade Estadual Paulista
cabrini@marilia.unesp.br

1 Introdução

Por refletir o grau de desenvolvimento de uma nação e proporcionar diretrizes nos âmbitos social, político e econômico, a avaliação da produção científica tornou-se um fator imprescindível, assim como, a elaboração de critérios que avaliem os diversos domínios científicos.

Entre as abordagens para a caracterização e avaliação de um domínio, destacam-se os estudos bibliométricos, especialmente as análises baseadas nas citações, por constituir abordagem consolidada e objetiva (HJØRLAND, 2002).

Nesse contexto, destacam-se os esforços para se elaborar mecanismos adequados de avaliação dos domínios científicos, assim como de se dispor de instrumentos e indicadores para definição de suas diretrizes, alocação de investimentos e recursos, formulação de programas e avaliação de atividades relacionadas ao desenvolvimento científico e tecnológico, como forma de identificar as elites científicas, garantindo um investimento profícuo das agências de fomento à pesquisa. (MUGNAINI *et al.,* 2004, THOMAZ *et al.*, 2011, VANZ; STUMPF, 2010).

Diante dessa necessidade, Hirsch (2005) desenvolveu um índice bibliométrico, denominado *índice-h,* para a avaliação do desempenho científico dos pesquisadores que considera simultaneamente aspectos relativos ao impacto (citações) e à produtividade (número de artigos) do pesquisador.

Este índice apresenta algumas características que devem ser consideradas durante a análise do desempenho dos pesquisadores: não deve ser utilizado para comparar pesquisadores de áreas diferentes e pode atribuir vantagens para aqueles que têm maior tempo de titulação.

Egghe (2010) observa que novos pesquisadores têm clara desvantagem quando comparados aos que possuem carreira mais longa. Além disso, o *índice h* pode aumentar, mesmo que não haja novos artigos publicados.

Esse autor desenvolveu uma variação de *h* denominada *índice g,* definida por "um conjunto de artigos tem *índice g* se *g* é o posto mais alto, de modo que os artigos principais têm juntos, pelo menos g^2 citações" (EGGHE, 2010). Uma das características deste indicador é atribuir maior peso ao número de citações recebidas pelo artigo. Todavia, também o índice h pode atribuir maiores valores para pesquisadores com maior tempo de publicação.

Nesse contexto, considera-se que a relativização desses indicadores em função do tempo de publicação do pesquisador pode eliminar a dependência do *índice h e índice g* em relação ao tempo (*t*), propiciando que se visualize de forma mais fidedigna o impacto e inserção do pesquisador na comunidade científica.

Para o índice h, o indicador relativo h/t é definido por: d*ado o índice h do autor e o tempo t de publicação, obtém-se a razão h/t, segundo a qual o autor publicou, em média, h/t artigos com pelo menos h/t citações cada, por ano.*

Para o índice g, o indicador relativo g/t é definido por: d*ado o índice g do autor e o tempo t de publicação, obtém-se a razão g/t,* segundo a qual *o* pesquisador *publicou, em média, g/t artigos* por ano*, os* quais têm, juntos, pelo menos g^2/t citações*.*

Considerando o exposto, esta pesquisa tem por objetivo realizar uma análise comparativa entre o índice h, índice g e suas forma relativizadas pelo tempo de publicação, dos investigadores mais produtivos em "Estudos Métricos", temática consignada à área de Ciência da Informação, a fim de se avaliar as tendências e adequação dos indicadores na descrição do impacto e inserção dos pesquisadores na comunidade científica.

De forma mais específica, busca-se identificar, descrever e analisar o índice h, índice g e os índices relativos h/t e g/t, tomando como universo os pesquisadores mais produtivos da revista *Scientometrics,* considerada representativa da temática Estudos Métricos na ciência *mainstream*, no período de 1980 a 2014.

Esta pesquisa se justifica pela contribuição para o aprofundamento da análise dos índices do tipo h de Hirsch e da relação entre os mesmos. Auxiliará na compreensão desses indicadores como descritores da trajetória acadêmica de um pesquisador na temática Estudos Métricos.

2 Procedimentos metodológicos

Para verificar o comportamento dos indicadores, tomou-se como universo, os 20 autores mais produtivos na revista *Scientometrics*, correspondendo a ter pelo menos 20 artigos científicos publicados, no período de 35 anos (1980-2014). No período de 6 a 10 de fevereiro de 2015, para cada pesquisador, levantaram-se, na base *Scopus*, os artigos publicados, o número de citações por artigo e o ano de publicação.

A partir dos dados levantados, calcularam-se os índices *h* e *g,* a razão *h/t* e *g/t* de cada investigador. Para cada indicador, obtiveram-se: média; desvio padrão; máximo; mínimo; coeficiente de variação e correlações de Pearson, a fim de verificar a tendência e a dispersão do grupo quanto aos indicadores estudados e a correlação entre os índices h e g absolutos e os relativos h/t e g/t.

3 Análise e interpretação dos resultados

A Tabela 1 apresenta os valores dos índices h e g e suas relativizações pelo tempo de publicação no periódico Scientometrics dos 20 investigadores mais produtivos na revista, ordenados em função do índice h dos pesquisadores.

Verifica-se que o autor com maior *índice h* é Glänzel (*h=36*). O autor vem publicando na *Scientometrics* há 31 anos. Em contrapartida, Gupta tem o menor índice (*h = 5*), publica na *Scientometrics* há 12 anos, um período relativamente menor quando comparado ao primeiro pesquisador.

O *índice h* médio dos pesquisadores está em torno de 15, ou seja, eles têm pelo menos 15 artigos com pelo menos 15 citações cada. Além disso, pode-se dizer que a dispersão do grupo é 7,4 artigos com pelo menos 7,4 citações cada.

Torna-se pertinente verificar o comportamento do *índice h* relativizado pelo período de tempo (*h/t*) que os investigadores publicam na temática “Estudos Métricos”, na Revista *Scientometrics*.

Glänzel publica na Scientometrics há 31 anos, seu *índice h* é 36 e a razão *h/t* de 1,16, ou seja, ele publicou, em média, 1,16 artigos com pelo menos 1,16 citações cada, por ano. Guan e Thelwall têm um índice *h/t* de 1, eles publicaram 1 artigo com pelo menos uma citação, por ano.

Tabela 1. Indicadores bibliométricos para os 20 autores mais produtivos na revista *Scientometrics* (1980-2014).

Pesquisador	Índice h	Índice g	Tempo t (anos)	Razão h/t	Razão g/t
Glänzel, W.	36	60	31	1,16	1,94
Schubert, A.	28	59	33	0,85	1,79
Braun, T.	21	40	33	0,64	1,21
Leydesdorff, L.	19	32	33	0,58	0,97
Van Raan, A.F.J.	19	36	26	0,73	1,38
Moed, H.F.	18	32	29	0,62	1,10
Meyer, M.	14	22	16	0,88	1,38
Rousseau, R.	14	25	27	0,52	0,93
Vinkler, P.	14	22	26	0,54	0,85
Lewison, G.	12	19	23	0,52	0,83
Thelwall, M.	12	22	12	1,00	1,83
Bar-Ilan, J.	12	19	17	0,71	1,12
Moya-Anegon, F.	11	19	16	0,69	1,19
Egghe, L.	11	32	28	0,39	1,14
Garg, K.C.	11	15	29	0,38	0,52
Guan, J.	10	16	10	1,00	1,60
Abramo, G.	8	14	6	1,33	2,33
D'Angelo, C.A.	8	14	6	1,33	2,33
Kretschmer, H.	8	16	31	0,26	0,52
Gupta, B.M.	5	6	12	0,42	0,50
Média	15	26,0	22,2	0,73	1,27
Desvio Padrão	7,4	14,2	9,3	0,31	0,55
Máximo	36	60	33	1,33	2,33
Mínimo	5	6	6	0,26	0,50
Coef.Variação	51%	55%	42%	42%	43%

Fonte: elaborado pelos autores.

Os investigadores Abramo e D’Angelo publicam há 6 anos, e têm *índice h* de 8 e *h/t=1,33*, ou seja, eles publicaram, em média, 1,33 artigos com pelo menos 1,33

citações cada, por ano. Pode-se notar que os autores publicam nesse periódico há poucos anos e, no entanto, esses artigos tiveram relevância e foi refletida no *índice h*.

As Figuras 1a e 1b permitem uma melhor visualização da dispersão dos índices relativos em função dos índices absolutos, h e g. Observa-se na Figura 1a que dos 20 pesquisadores analisados, cinco (25%) dos pesquisadores apresentaram tendências bastante distintas entre índice h e seu derivado, relativizado pelo tempo de publicação: D'Angelo, Abramo, Guan, Thelwall e Meyer apresentaram menores índices h e altos índices h/t, por outro lado Braun apresentou alto índice h e menor valor para h/t.

Figura 1a. Gráfico de dispersão do índice relativo h/t em função do índice h

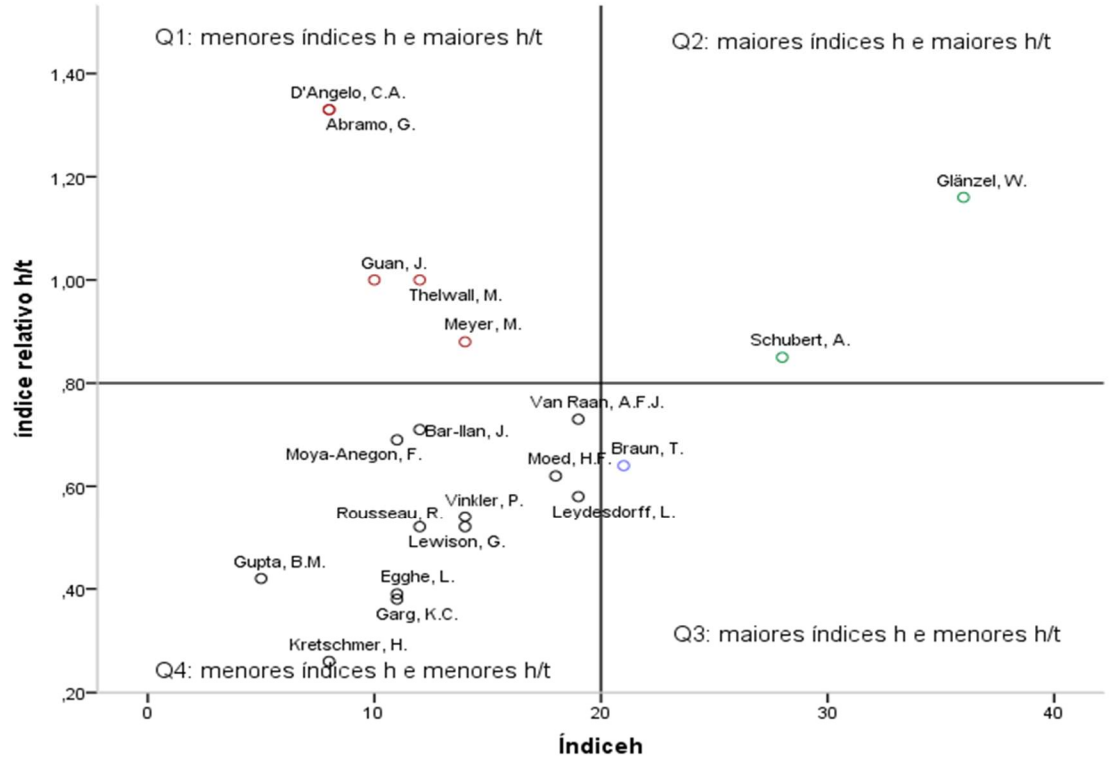
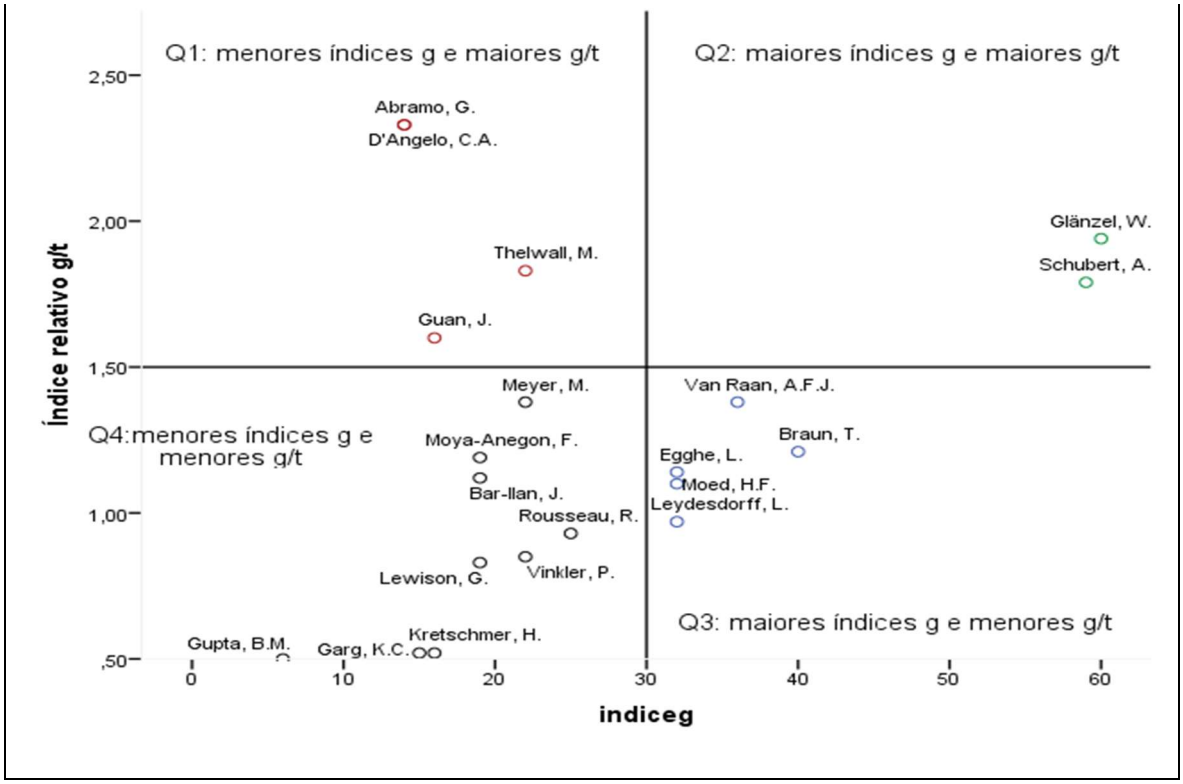


Figura 1b. Gráfico de dispersão do índice relativo g/t em função do índice g



Fonte: elaborado pelo autores.

Ainda, observam-se autores que publicam há muitos anos na revista e, entretanto, o índice médio é relativamente baixo, tais como Kretschemer, Garg, e Egghe. A grande maioria dos autores analisados publica na *Scientometrics*, na temática estudada, em média, um artigo a cada dois anos.

Ao considerar o índice relativo *h/t*, os autores mais recentes se equiparam com os mais antigos em tempo de publicação, por eliminar a dependência do tempo de publicação. Este indicador relativo supera a desvantagem apresentada pelo *índice h*, em que pesquisadores com maior tempo de titulação têm maiores vantagens quando comparado a pesquisadores mais recentes.

O valor de correlação de Pearson entre estes dois índices, igual a 0,203, evidencia que a tendência dos valores desses dois indicadores não apresenta associação estatisticamente significante, indicado que representam características distintas dos pesquisadores.

Em relação ao *índice g*, identifica-se, similarmente, que Glänzel tem o maior *índice g* e Gupta o menor (*g = 6*). O *índice g* médio dos autores é 26, significando: este é o maior número, nos quais os artigos mais citados receberam conjuntamente, pelo menos 676 citações (26^2 citações). Destaca-se que a variabilidade do grupo é praticamente o dobro da observada para o *índice h*.

A razão *g/t* analisa o quociente entre o *índice g* e o tempo de publicação no periódico. Em média, os investigadores têm um quociente de 1,27, este é o maior número nos quais os artigos mais citados receberam juntos, pelo menos 1,6 citações ($1,27^2$ citações), por ano.

Verifica-se, pela Figura 1b, que Abramo, D'Angelo, Thelwall e Guan têm altos valores para os índices relativizados (g/t) e menores valores de g, evidenciando tendências distintas em relação a estes dois índices.

Assim como para os índices h e h/t, o valor de correlação de Pearson entre g e g/t, igual a 0,288, evidencia que não há associação estatisticamente significante entre eles.

Ao analisar os indicadores relativos *h/t* e *g/t* identificam-se, praticamente os mesmos autores com produtividade e impacto elevados, sendo assim, essas métricas também auxiliam a análise métrica de pesquisadores em uma temática.

4 Considerações finais

Verifica-se que os índices *h* e *g* retratam bem o desempenho científico de investigadores na temática Estudos Métricos. Observa-se que indicador *h* não valoriza artigos altamente citados e nem aqueles com poucas citações. Em contrapartida, o *índice g* atribui maior peso ao número de citações que o artigo recebeu. Sendo assim, considera-se que os indicadores conseguem identificar aspectos distintos do impacto do pesquisador e são complementares em uma possível avaliação.

As razões *h/t* e *g/t* podem complementar a avaliação de um autor quando são analisados pesquisadores com diferentes tempos de publicação, visto que esse quociente elimina a influência do tempo. Desse modo, a primeira relação determina o *índice h* por ano e a segunda, determina as citações recebidas conjuntamente pelos artigos publicados, por ano.

Finalizando, considera-se que para avaliar fidedignamente o desempenho científico de um pesquisador, temática, disciplina, área do conhecimento ou país é indicado utilizar uma combinação desses indicadores na análise realizada.

Referências

EGGHE, L. **The Hirsch-Index and Related Impact Measures**, 2010. Disponível em: https://uhdspace.uhasselt.be/dspace/bitstream/1942/10289/2/Measures%202.pdf. Acesso em: 15 set 2014.

HIRSCH, J. E. **An index to quantify an individual's scientific research output**, 2005.

HJØRLAND, B. Domain analysis in information science: eleven approaches traditional
as well as innovative**. Journal of Documentation**, v. 58, n. 4, p.422-462, 2002

MUGNAINI, R.; JANNUZZI, P. M.; QUONIAM, L. **Indicadores bibliométricos da produção científica brasileira**: uma análise a partir da base Pascal. Revista Ciência e Informação. v. 33, p. 123-131, 2004.

THOMAZ, P. G; ASSAD, R. S.; MOREIRA, L. F. P. **Uso do fator de impacto e do índice h para avaliar pesquisadores e publicações**. Arquivo Brasileiro de Cardiologia, p. 90-93, 2011.

VANZ, S. A. S.; STUMPF, I. R. C. **Procedimentos e ferramentas aplicados aos estudos bibliométricos**. Informação e Sociedade, v. 20, n.2, p. 67-75, maio/ago. 2010.

A abordagem da análise de domínio na organização e representação do conhecimento em arquivística

Cynthia Maria Kiyonaga Suenaga
Universidade Estadual Paulista
cmksuenaga@gmail.com

Brígida Maria Nogueira Cervantes
Universidade Estadual de Londrina
brigidacervantes@gmail.com

1 Introdução

A preocupação com uma possível análise de informações documentais, visando o conhecimento, é o que conduz, nesta pesquisa, à busca de metodologias para se estruturar e ordenar o conhecimento armazenado na forma de documentos individuais ou de conjuntos documentais no interior dos arquivos.

A perspectiva levantada pelo paradigma científico-informacional (voltado para o conteúdo informacional registrado nos documentos), em detrimento ao paradigma documental (preocupado com a custódia e preservação) observa o documento não mais como um objeto físico a ser armazenado e preservado apenas e sim utilizado e analisado de diversas formas, dando início a uma nova forma de pensar os arquivos.

A Arquivística é uma disciplina do conhecimento humano que necessita de trabalhos de pesquisa para direcionar suas aplicações, de contínua reflexão sobre os modelos de interpretação e de intervenção na realidade (SOUSA, 2012). Proposta por Hjørland e Albrechtsen (1995) como um novo enfoque para o estudo do campo da Documentação e da Ciência da Informação (CI) a Análise de Domínio constitui-se em uma formulação teórico-metodológica que por meio da união de teorias de base e procedimentos aplicados provenientes da CI permite a Organização do Conhecimento nesta e em outras áreas.

O Plano de Classificação de Documentos (PCD), sob o aspecto do conteúdo informacional dos documentos e das funções e atividades que lhes deram origem, pode contribuir para a identificação e entendimento das estruturas de conhecimento existentes no arquivo. Segundo Héon (1995) o PCD é o primeiro instrumento de rastreamento intelectual para documentos em qualquer nível. Desta forma, baseando-se em uma das prerrogativas para a elaboração do PCD que é uma cuidadosa avaliação da instituição geradora ou mantenedora dos documentos, seleciona-se o PCD como o objeto de pesquisa a ser analisado na busca de identificar o conhecimento existente no Domínio de um arquivo.

A partir do exame de um dos principais produtos do tratamento documental em arquivos questiona-se como a abordagem da Análise de Domínio no contexto da Organização do Conhecimento em CI pode contribuir para a compreensão da estrutura de conhecimento presente em Arquivos Públicos Estaduais Brasileiros?

O objetivo desta pesquisa foi investigar estruturas de conhecimento presentes em Arquivos Públicos Estaduais Brasileiros (APEBs), sob a perspectiva da Análise de Domínio. Com esse intuito buscou-se identificar a existência de Planos de Classificação de Documentos em Arquivos Públicos Estaduais Brasileiros; verificar a presença de normas para a elaboração do Plano de Classificação de Documentos; relacionar os Planos de Classificação de Documentos ao Domínio e à comunidade discursiva para selecionar os conceitos e possíveis relacionamentos presentes em Plano de Classificação de Documentos.

2 Pressupostos Teóricos

Uma característica essencial da informação-como-conhecimento (BUCKLAND, 1991), perspectiva que adotamos nesta pesquisa, é sua forma intangível, portanto para expressá-la é preciso descrevê-la ou representá-la e neste momento torna-se informação-como-coisa, algo que se pode manusear de alguma forma. Sob este enfoque, representações de conhecimento formam subconjuntos de informação-como-coisa, as quais podem se materializar como documentos ou como Sistemas de Organização do conhecimento, que são importantes caminhos para se lidar com a informação-como-conhecimento. Os Sistemas de Organização do Conhecimento, produto da Organização do Conhecimento, registram aspectos do documento que

estão ligados à questões relacionadas diretamente com o Domínio e com aspectos sócio-histórico-culturais que são evidenciados pela Análise de Domínio.

A informação é agente mediador na produção de conhecimento e aparece em forma e conteúdo, como estruturas significantes, capazes de gerar conhecimento individual e coletivo (BARRETO, 1994). Tais estruturas são constituídas por convenções e acordos entre indivíduos que convivem em grupos nos limites de seu domínio.

De acordo com Hjørland e Albrechtsen (1995), tais grupos são as comunidades discursivas as quais possuem um papel na sociedade que se reflete na organização do conhecimento, na estrutura informacional, nas formas de linguagem e comunicação e nos sistemas de informação, padrões de cooperação e critérios de relevância. Para Swales (1990) comunidade de discurso é um grupo de pessoas que se ligam em primeiro lugar para perseguir objetivos comuns, as necessidades comunicativas voltadas para o cumprimento de metas ou objetivos tendem a predominar no desenvolvimento e manutenção de suas características.

Hjørland e Albrechtsen (1995) baseiam a teoria da Análise de Domínio na constatação de que o Domínio é constituído pela comunidade discursiva e que o registro de seus processos comunicativos reflete o conhecimento circulante neste ambiente. Mais tarde Hjørland (2002) apresenta onze formas de abordar-se um Domínio visando elucidar esse conhecimento, as referidas abordagens englobam fazeres fundamentais da CI tais como indexação, mapas bibliométricos, estruturas de Comunicação Científica, estudos históricos, epistemológicos, terminológicos e críticos, o autor ainda afirma que outras podem surgir e que é aconselhável o uso de dois ou mais enfoques concomitantemente.

A seguir apresenta-se uma representação das linhas gerais que fundamentam a abordagem da Análise de Domínio em CI (Figura 1).

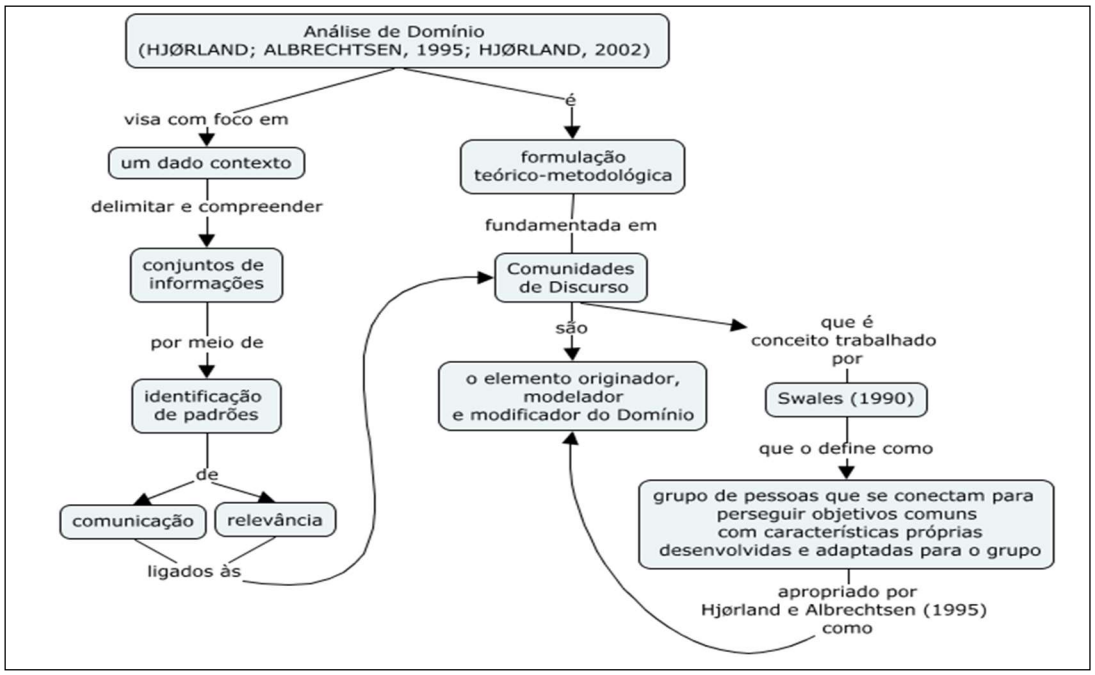


Figura 1 – Linhas gerais da Análise de Domínio

Fonte: Elaborado pelo autor.

O foco da Análise de Domínio é delimitar e compreender o conjunto de informações em um dado domínio a partir da comunidade de discurso que o constitui e o modifica no espaço e no tempo. Conforme Foucault (2008) o discurso não é simplesmente aquilo que manifesta ou oculta o desejo, mas também aquilo que é objeto do desejo; o discurso não é simplesmente aquilo que traduz as lutas ou os sistemas de dominação, mas aquilo por que se luta. A fundamentação desta abordagem na comunidade discursiva amplia o horizonte de pesquisa do Cientista da Informação, uma vez que olhar para o discurso envolve sempre variáveis históricas, contextuais e sociais.

Tennis (2003) busca aproximar a teoria da Análise de Domínio da realização concreta de estudos informacionais voltados para o Domínio apresentando o que se pode chamar de desenvolvimento da Análise de Domínio, propondo dois eixos que consistem em dispositivos analíticos que complementam a AD. Segundo o autor embora as onze abordagens para a Análise de Domínio propostas por Hjørland (2002) ofereçam subsídios para estudar o domínio não oferecem definição estruturada e delimitação sistematizada daquilo que se está analisando.

A área de modulação é o eixo no qual por meio de um enunciado, nomeia-se e delimita-se a extensão do domínio examinado, indicando o que está incluso ou não e

como o domínio será chamado (definem-se termos pertinentes e a extensão do Domínio). E o grau de especialização, qualifica e marca a intenção do Domínio, diminuindo sua extensão e aumentando sua intenção, o qualificador do domínio é o foco dado à análise (TENNIS, 2003).

Conforme NORDBERG, LINELL e GUNARSON (1997) eventos falados tais como telefonemas, reuniões, conferências, entre outros são intermediados, precedidos e seguidos de práticas escritas em forma de documentos como memorandos, relatórios, contratos e outros. Os autores consideram que a comunicação efetiva em discurso é escrita e fundamental para a sociedade funcionar. Este discurso ocorre não somente entre o grupo profissional, mas também entre diferentes grupos que podem ser considerados como comunidades discursivas.

A materialidade dos discursos da comunidade discursiva concretiza-se pela institucionalização e socialização do discurso, na sua formalização em textos, em seu sentido amplo, incluindo escrita, visual, audiovisual. O discurso documental se materializa em textos documentais e o resultado de seu estudo pode ser visto como um saber científico que se constitui na relação com as instituições e com os sujeitos da ciência, assim como os meios de circulação do saber (SALES, 2011; NUNES, 2008). Os textos, nessa direção, são discursos materializados, institucionalizados e socializados por meio dos documentos.

Dessa forma torna-se desejável buscar formas de identificar o discurso documental, para entender a formação de uma comunidade discursiva e para buscar elucidar o conhecimento registrado nos diversos domínios.

3 Procedimentos metodológicos

Realizou-se um levantamento preliminar dos Arquivos Públicos Estaduais Brasileiros com base em informações constantes no *website* do Conselho Nacional de Arquivos (CONARQ), que disponibiliza uma listagem com os Arquivos Públicos Estaduais Brasileiros. Na referida listagem identificou-se 24 Arquivos Públicos Estaduais dos quais 12 apresentavam endereço para o *website*. Esta primeira seleção passou por uma etapa de conferência na qual se verificou que algumas URLs não estavam acessíveis. Em virtude dessas inconsistências identificadas durante a

primeira tentativa de acesso realizou-se uma busca pelos demais *websites* dos APEBs, sem URL na listagem (ver Figura. 2).

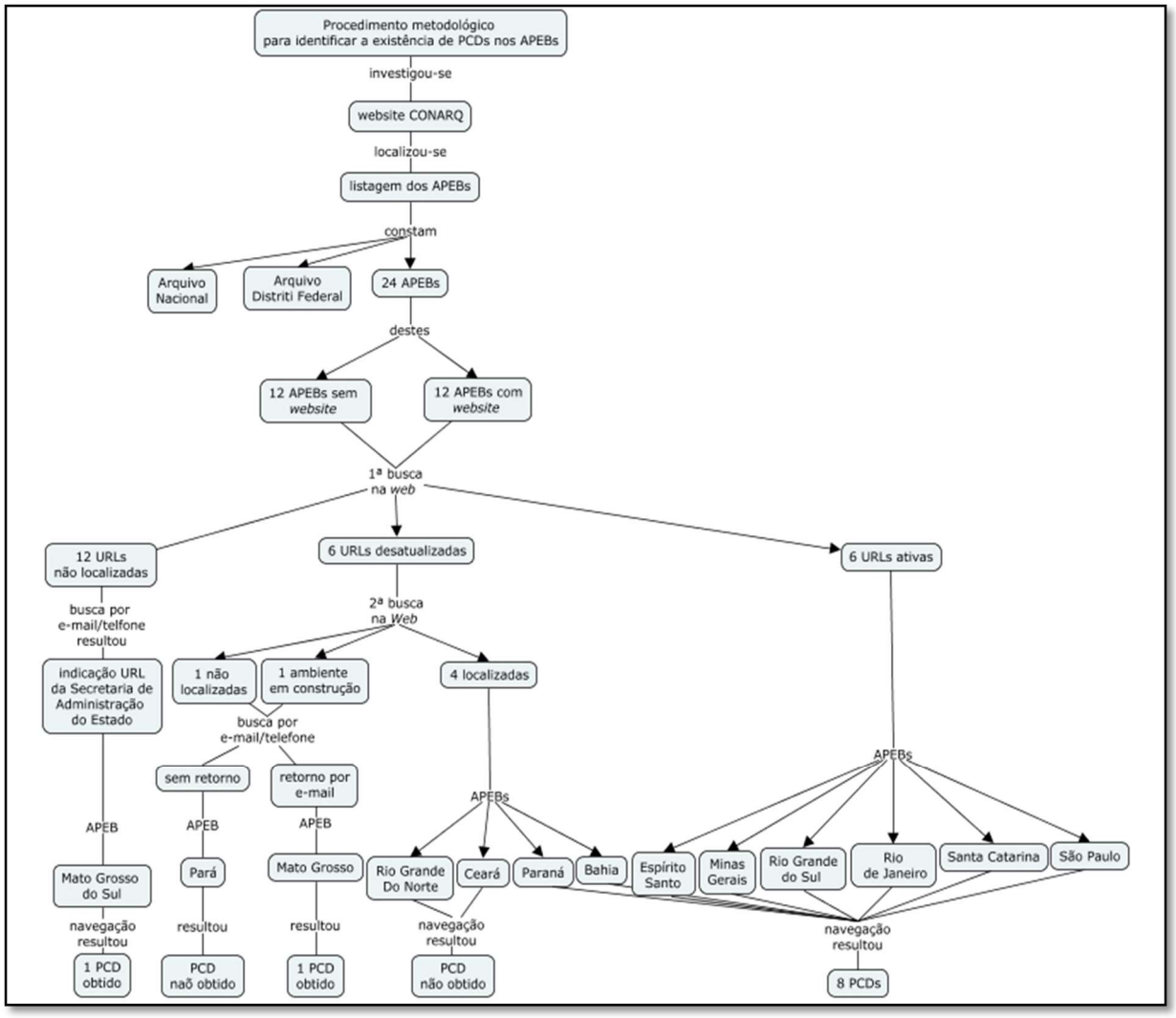


Figura 2 – Esquema da busca pelos PCDs

Fonte: Elaborado pelo autor.

Utilizando-se das 10 URLs reunidas entre o site do CONARQ e a *Web* realizou-se, então, a navegação nos *websites* em busca dos PCDs. A etapa seguinte foi a leitura e análise dos PCDs obtidos.

A partir da leitura dos PCDs reuniu-se as informações consideradas mais relevantes para a compreensão dos PCDs tais como a nomenclatura do PCDs, a descrição do escopo nas Classes principais, a descrição do escopo nos subníveis, o tipo de categorização e as classes principais (Figura. 3) .

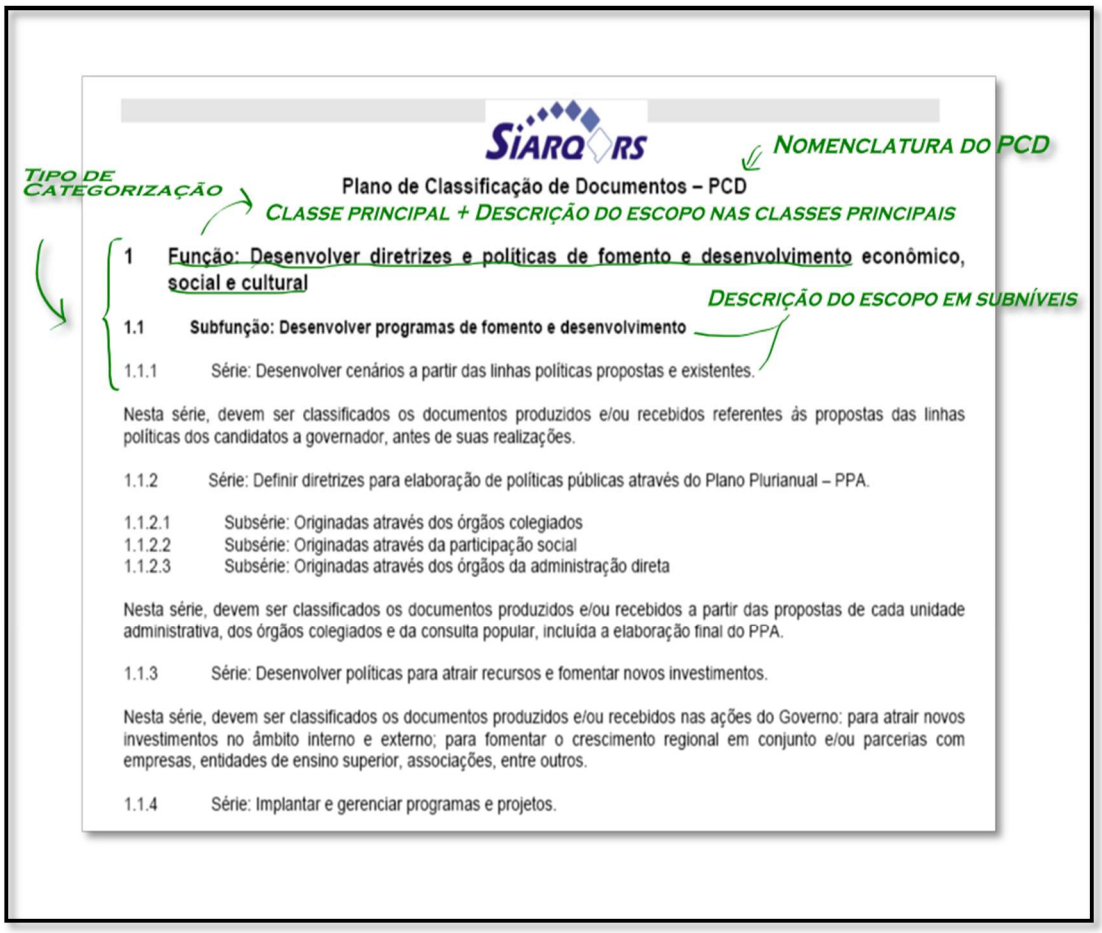

SIARQ RS
NOMENCLATURA DO PCD
TIPO DE CATEGORIZAÇÃO
Plano de Classificação de Documentos – PCD
CLASSE PRINCIPAL + DESCRIÇÃO DO ESCOPO NAS CLASSES PRINCIPAIS
1 Função: Desenvolver diretrizes e políticas de fomento e desenvolvimento econômico, social e cultural
DESCRIÇÃO DO ESCOPO EM SUBNÍVEIS
1.1 Subfunção: Desenvolver programas de fomento e desenvolvimento
1.1.1 Série: Desenvolver cenários a partir das linhas políticas propostas e existentes.
Nesta série, devem ser classificados os documentos produzidos e/ou recebidos referentes às propostas das linhas políticas dos candidatos a governador, antes de suas realizações.
1.1.2 Série: Definir diretrizes para elaboração de políticas públicas através do Plano Plurianual – PPA.
1.1.2.1 Subsérie: Originadas através dos órgãos colegiados
1.1.2.2 Subsérie: Originadas através da participação social
1.1.2.3 Subsérie: Originadas através dos órgãos da administração direta
Nesta série, devem ser classificados os documentos produzidos e/ou recebidos a partir das propostas de cada unidade administrativa, dos órgãos colegiados e da consulta popular, incluída a elaboração final do PPA.
1.1.3 Série: Desenvolver políticas para atrair recursos e fomentar novos investimentos.
Nesta série, devem ser classificados os documentos produzidos e/ou recebidos nas ações do Governo: para atrair novos investimentos no âmbito interno e externo; para fomentar o crescimento regional em conjunto e/ou parcerias com empresas, entidades de ensino superior, associações, entre outros.
1.1.4 Série: Implantar e gerenciar programas e projetos.

Figura 3 – Exemplo das informações levantadas nos PCDs

Fonte: Elaborado pelo autor.

Essas informações são posteriormente analisadas em aproximação com as teorizações de Tennis (2003) e HJØRLAND (2002).

4 Análise e resultados

Os PCDs dos APEBs, por meio dos títulos dados ao documento, apresentam de alguma forma informações pertinentes ao eixo área de modulação de Tennis (2003) neste sentido quanto mais específico for o título, menor será a extensão do Domínio e maior a sua intenção. Ou seja, mais nítido se torna o objetivo ou propósito do Domínio.

No caso dos PCDs entende-se que o título do documento possui a função de especificar o Domínio ao qual se aplica a classificação, isto é esclarecer da maneira mais eficiente possível quais documentos classificam-se em cada nível. No exemplo dos PCDs a maior parte deles especifica no título do documento os limites: Estadual do/de (nome do Estado), do Poder Executivo, das Atividades-Meio.

A leitura dos PCDs dos APEBs demonstra que em sua grande maioria apresentam descrição do escopo nos níveis e subníveis de categorização. Identifica-se, no entanto, que essa descrição, quando existente, não é criteriosa em todos os casos. Algumas são minuciosas, tais como as dos APEBs da Bahia, e São Paulo. Outras são mais genéricas não dando muitos indícios a respeito da extensão e intenção do domínio.

A forma de identificação do discurso se deu por meio da identificação dos conceitos expressos na nomenclatura apresentada nas entradas do PCD, as categorias principais (entradas), aqui compreendidas conforme Dahlberg (1978) como conceitos em sua mais ampla extensão dão indicações da extensão do subdomínio.

APE-BA	APEMS	APESP
Organização Administrativa	1 Organização Administrativa	01 Organização Administrativa
Comunicação Institucional	2 Comunicação Institucional	02 Comunicação Institucional
Gestão de Recursos Humanos	3 Gestão de Recursos Humanos	03 Gestão de Recursos Humanos
Gestão de Bens Materiais e Patrimoniais	4 Gestão de Bens Materiais e Patrimoniais	04 Gestão de Bens Materiais e Patrimoniais
Gestão Orçamentária	5 Gestão Orçamentária e Financeira	05 Gestão Orçamentária e Financeira
Gestão de Documentos e Informações	6 Gestão de Documentos e Informações	06 Gestão de Documentos e Informações
Gestão de Atividades Complementares	7 Gestão de Atividades Complementares	07 Gestão de Atividades Complementares

Quadro 1 – Identificação do Discurso das categorias principais dos APEBs

Fonte: Elaborado pelo autor.

A semelhança entre o discurso das categorias principais elencadas nos PCDs dos APEBs relacionados (ver Quadro 1) e também a identificação das características das comunidades discursivas de Swales (1990) associando-as aos APEBS (ver

Quadro 2) indicam a constituição de uma Comunidade Discursiva. O uso de praticamente os mesmos termos demonstra a proximidade do discurso e indica que o objetivo da comunidade discursiva é partilhado.

Característica	**Identificação nos APEBs**
Conjunto de objetivos públicos comuns de amplo acordo	Conforme a lei de Arquivos: Conforme a lei de Arquivos: "[...] a gestão documental e a proteção especial a documentos de arquivos, como instrumento de apoio à administração, à cultura, ao desenvolvimento científico e como elementos de prova e informação." (BRASIL, 1991).
Mecanismos de intercomunicação entre seus membros	Uso dos mecanismos usuais de comunicação institucional tais como e-mail, ofícios, memorandos, entre outros.
Uso dos mecanismos participativos para troca de informação e *feedback*	Demonstrado por exemplo pela constituição de Comissões de Avaliação de Documentos.
Desenvolvimento de conjuntos de gêneros	Demonstrado pela produção de PCDs, Tabelas de Temporalidade de Documentos, Termo de Descarte de Documentos, entre outros
Aquisição de léxicos específicos	Uso de Siglas tais como TTD, SIGAD, entre outras e uso de termos técnicos.
Um limiar de membros com grau adequado de conteúdo relevante e expertise discursiva	Presença obrigatória de Arquivistas entre os membros do grupo de trabalho.

Quadro 2 – Identificação dos APEBs como Comunidade Discursiva

Fonte: Elaborado pelo autor.

O Discurso das categorias principais nos PCDs dos APEBs do Espírito Santo, Mato Grosso, Minas Gerais, Paraná, Rio de Janeiro e Santa Catarina não apresentam a mesma proximidade por semelhança que os APEBs da Bahia, Mato Grosso do Sul e São Paulo, mas existem semelhanças nítidas considerando-se os conceitos abordados. A terminologia utilizada para expressar as categorias não é a mesma, mas o conceito geral expresso é o mesmo ou muito semelhante.

5 Considerações finais

A caracterização dos APEBs e consequentemente de outras instituições como Comunidades Discursivas traz aos pesquisadores e profissionais um embasamento teórico-metodológico que é bem vindo à área de Organização do Conhecimento e especialmente à Arquivística, a realização de mais pesquisas voltadas para o uso da Análise de Domínio nestes campos de estudo é recomendada, visando promover o seu desenvolvimento.

O que se observa em linhas gerais é que elementos de uma Comunidade Discursiva produzem discursos homogêneos ainda que não idênticos. Um aumento na capacidade de troca de informações entre os membros do Domínio certamente aumentaria o nível de *expertise* geral e de novas soluções para o Domínio.

Almeja-se a realização de estudos que auxiliem no delineamento do conhecimento existente no Domínio da Arquivística, no qual os Documentos são objeto de estudo, assim como seu conteúdo, desde a gênese da acumulação de documentos que dá origem ao Arquivo enfatizando o discurso, em busca do entendimento daquilo que essa acumulação comunica. Pesquisas desse tipo abrem caminho para a Organização de Conhecimento estabelecer-se como ferramental essencial para o Arquivista.

Referências

DAHLBERG, I. Teoria do Conceito. **Ciência da Informação**, Rio de Janeiro, v. 7, n.2, p. 101–107, 1978b. Disponível em: <http://revista.ibict.br/ciinf/index.php/ciinf/issue/view/87/showToc>. Acesso em: 10 maio 2013.

BARITÉ, M. Organización del conocimiento: um nuevo marco teórico-conceptual em bibliotecología y documentación. In: CARRARA, K. (Org.). **Educação Universidade e Pesquisa**. Marília: Unesp - Marília, 2001. p. 35-60.

BARRETO, A. A questão da informação. **São Paulo em Perspectiva**: revista da Fundação SEADE, v. 8, n. 4, Out-Dez, p. 3-8, 1994. Disponível em: <http://www.seade.gov.br/produtos/spp/index.php?men=res>. Acesso em: 20 out. 2012.

BUCKLAND, M. K. Information as thing. Journal of American Society for Information Science. V. 5., n. 42, p. 351-360, 1991. Disponível em: <http://people.ischool.berkeley.edu/~buckland/thing.html>. Acesso em: 15 jun. 2012.

FOUCAULT, M. **A ordem do discurso**. São Paulo: Loyola, 2008.

HÉON, G. Les plans de classification en archivistique. **Archives**, v. 27, n. 1, 1995. Disponível em: <http://www.archivistes.qc.ca/revuearchives/vol27_1/27-1-heon.pdf>. Acesso em: 16 fev. 2014.

HJØRLAND, B; ALBRECHTSEN, Hanne. Toward a New Horizon in Information Science: Domain-Analysis. **Journal of The American Society for Information Science**: JASIS**,** v.46, n.6, 400-425, 1995. Disponível em: <http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199507)46:>. Acesso em: 14 jul. 2012.

HJØRLAND, B. Domain analysis in information science: eleven approaches – traditional as well as innovative. **Journal of Documentation**, v.58, n.4, 422-462, 2002. Disponível em:<http://dx.doi.org.ez78.periodicos.capes.gov.br/10.1108/00220410210431136>. Acesso em: 16 jun. 2012.

NORDBERG, B.; LINELL P.; GUNNARSSON B. (Eds).**The construction of Professional Discourse**. London: Longman, 1997.

NUNES, J. H. O discurso documental na história das ideias linguísticas e o caso dos dicionários. **Alfa**: Revista de Linguística, São Paulo, v. 52, n. 1, p. 81-100, 2008. Disponível em: <http://seer.fclar.unesp.br/alfa/article/view/1468/0>. Acesso em: 28 set. 2012.

SALES, R. de. Representação de domínios em Biblioteconomia e Arquivística. In: SILVA, F. C. C. da; SALES, R. de. (Orgs.) **Cenários da Organização do Conhecimento**: linguagens documentárias em cena. Brasília: Thesaurus, 2011.

SOUSA, R. T. B. de. A classificação como função matricial do que fazer arquivístico. In: SANTOS, V. B. dos; INNARELLI, H. C.; SOUSA, R. T. (Orgs.). **Arquivistica:** temas contemporâneos: classificação, preservação digital, gestão do conhecimento. 38 ed: Distrito Federal: SENAC, 2012.

SWALES, J. M. The concept of discourse community. In: SWALES, J. M **Genre analysis:** English in Academic and Research Settings. Cambridge: Cambridge University Press, 1990. p. 21–27.

TENNIS, J. T. Two Axes of Domains for Domain Analysis. **Knowledge Organization**, v. 3/4, jul. 2003. Disponível em: <http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=14887661&lang=pt-br&site=ehost-liv>. Acesso em: 20 ago. 2013.

O *Familysearch Indexing* como iniciativa de *crowdsourcing* no contexto da organização do conhecimento

Paula Carina de Araújo
Universidade Estadual Paulista
paula.carina.a@gmail.com

José Augusto Chaves Guimarães
Universidade Estadual Paulista
guima@marilia.unesp.br

1 Introdução

Na Ciência da Informação (CI), há muito se fala da importância de compartilhar conhecimentos diversos para gerar novo conhecimento. É o que Wersig (1993) chamou de "despersonalização do conhecimento" ao afirmar que as tecnologias de informação e comunicação surgiram para proporcionar cada vez mais a transmissão do conhecimento, independente da pessoa que o possui. Percebe-se aí o desenvolvimento de uma visão sociocognitiva, onde não somente o indivíduo é considerado, mas, o meio em que está inserido, suas relações e influências.

O *crowdsourcing* insere-se nesse contexto. Esse termo foi apresentado pela primeira vez por Howe (2006), no artigo *"The rise of crowdsourcing"*, no qual o autor indica exemplos de organizações que reconheceram na colaboração em massa uma oportunidade para acessar e produzir conteúdo barato, relevante e muitas vezes de qualidade, produzido por pessoas do mundo inteiro e disponível na internet.

A CI inserida na pós-modernidade é caracterizada pelo seu pluralismo de métodos, técnicas e reflexões - o que ocasiona um maior questionamento e crítica da área – e indica que os seus objetos começam a ser levados em consideração, conforme afirmam Tálamo e Smit (2007).

Hjorland (2002a) nessa mesma linha demonstrou preocupação por fortalecer o campo ao apresentar onze abordagens para o entendimento da CI e sua relação com

outros domínios. Nesta pesquisa, por exemplo, é proposto o relacionamento entre Organização do Conhecimento e a História da Família, tomando como objeto o *FamilySearch Indexing (FSI)*, “projeto que possibilita a indexação de milhões de registros genealógicos por milhões de voluntários, o que faz dele uma das maiores iniciativas de *crowdsourcing* no mundo”. (HANSEN et al. 2012, tradução nossa).

Dessa forma, apresenta-se como objetivo geral descrever o FSI no contexto da Organização do Conhecimento (OC). Tálamo e Smit (2007) sugerem que um possível programa de trabalho no domínio da informação deve contemplar cinco elementos, dentre eles, pode-se destacar dois que justificam essa pesquisa: - a articulação entre os dispositivos tecnológicos da informação, a produção da informação e a geração de sentido; - a inserção social da informação, com determinação de condições locais de recepção, visando o aperfeiçoamento dos dispositivos e o estudo da atividade dos usuários-consumidores.

2 Familysearch

O FamilySearch foi criado em 1989 como um sistema que abrangia todas as iniciativas anteriores de automação dos registros genealógicos reunidos desde a criação da Sociedade Genealógica de Utah em 1894. (ALLEN; EMBRY; MEHR, 1995).

Essa sociedade é uma instituição sem fins lucrativos, que presta atendimento por meio de sua principal biblioteca, a *Family History Library* localizada em *Salt Lake City, Utah*. Conta também com aproximadamente 4.000 centros de história da família, bibliotecas setoriais em 88 países e um website chamado de *FamilySearch* (www.familysearch.org) . Tem como missão “preservar e prover acesso à informação história que pode ser usada para reconhecer ancestrais e famílias, bem como para compilar genealogias ou histórias de famílias”. (METCALFE; THATCHER, 2008, p. 1).

“Antes da web, o *FamilySearch* utilizava formatos analógicos (disquetes, CDs, correio, etc.) para transmitir informação entre a biblioteca e os voluntários” (ELLIS, 2014, p. 7, tradução nossa). Considerado um exemplo de colaboração na web, como iniciativa de *crowdsourcing,* o FamilySearch foi muito beneficiado com o surgimento da internet.

Pelas características citadas anteriormente, o *FamilySearch* incentiva a participação de voluntários para a melhoria de seus produtos e serviços, bem como

para a disponibilização mais rápida dos registros de que dispõe. Podem ser citados alguns desses produtos e serviços: a nova Árvore Familiar; uma expressiva base de dados genealógica que inclui dados de projetos de indexação de registros e informação que provém da contribuição de pessoas do mundo todo, que não necessariamente são membros da Igreja (KENNARD; KENT; BARRETT, 2011); acesso a registros genealógicos em formato de microfilme, a outros que já foram digitalizados e a registros genealógicos digitalizados e indexados, o que os torna recuperáveis por meio de um sistema de busca e; indexação de registros por meio do *FSI,* tema deste estudo.

O FSI foi criado por iniciativa de A Igreja de Jesus Cristo dos Santos dos Últimos Dias, tendo em vista a valorização de sua doutrina para a preservação da família e a necessidade de conhecer os antepassados. Para colaborar com o projeto, é necessário criar uma conta na página no *FamilySearch,* fazer o download do programa FSI (uma versão web já vem sendo desenvolvida) no computador e acessá-lo com o *login* e senha. Posteriormente, é necessário escolher um lote de registros compatível com seu nível como indexador (iniciante, intermediário, avançado) e baixá-lo no computador. Esses lotes são conjuntos de registros genalógicos, podem ser: registros de nascimento, casamento, óbito, migração, etc. (IGREJA..., 2012).

Flinders (2012) explica que a “indexação é um processo de transcrição dos campos chave em um registro ou imagem digital. Não é uma transcrição completa, mas uma captura dos elementos chave do registro para que ele se torne recuperável”.

Hansen et al. (2013) afirmam que, desde sua criação, o *FSI* tem se dedicado para criar transcrições de alta qualidade. Para assegurar uma indexação correta, o projeto tem usado um processo para assegurar a qualidade baseado em um tipo de arbitragem, onde duas pessoas (A e B), independentemente, transcrevem as informações de uma imagem (nome, gênero e idade, por exemplo) e qualquer discrepância entre seu trabalho é avaliada por um árbitro experiente (ARB) que toma a decisão final. Esse modelo é chamado de A-B-ARB.

3 Familysearch indexing (FSI) e a abordagem sociocognitiva da organização do conhecimento (OC)

Ao estudar o uso do conhecimento pós-moderno relacionado à CI, Wersig (1993) preocupa-se com o a mudança do papel do conhecimento para os indivíduos, organizações e culturas e, para o autor, essa mudança envolve a dimensão filosófica e tecnológica. Nessa perspectiva, são os precursores do estudo paradigma social na CI Shera, Frohmann, Brier, Capurro e, com destaque, Hjorland.

É proposta a relação entre o FSI e a abordagem sociocognitiva por entender, assim como Barite (2001) que a OC procura englobar diversas práticas e atividades sociais relacionadas ao acesso ao conhecimento e busca agir como instrumento de tratamento da informação e de gestão do uso da informação de forma a integrar os fenômenos e as aplicações vinculadas à estruturação, disposição, acesso e difusão do conhecimento socializado.

No contexto da OC, a análise de domínio foi apresentada como “uma abordagem teórica da Ciência da Informação (CI), que afirma que a melhor maneira de compreender as informações na CI é o estudo de domínios por meio das comunidades discursivas”. (HJORLAND; ALBRECHTSEN, 1995, tradução nossa). Entende-se que está implícida na análise de domínio a abordagem sociocognitiva da OC.

Hjorland (2002b) explica que o ponto central em sua abordagem é que ferramentas, conceitos, significados, estruturas de informação, necessidades de informação e critérios de relevância são estabelecidos em comunidades discursivas onde o processo de comunicação é estabelecido. Percebe-se que nesta visão há uma mudança do foco da Ciência da Informação, dos indivíduos e/ou computadores para o mundo social cultural e científico.

Quando Hjorland e Albrechtesn (1995) afirmam que é necessário incorporar o conhecimento sobre as culturas nas quais os sistemas de informação estão funcionando é possível considerar o comportamento humano, as relações entre os participantes e o processo de indexação no FSI sob uma perspectiva sociocognitiva no domínio da OC.

Isso fica ainda mais evidente pela afirmação de Hjorland (2002b, p. 268) de que “o problema principal da CI é como as pessoas interpretam os textos para serem organizados e buscados assim como as necessidades de informação que devem ser satisfeitas”. Essa abordagem entende o conhecimento individual em uma perspectiva histórica, cultural e social.

Frohmann (2008, p.20-21) argumenta que o conceito de materialidade da informação “é muito importante quando se deseja investigar o que fazem os sistemas de informação [...] em sentido amplo para incluir o que ele chama de regimes de informação”. Ele defende a abordagem sociocognitiva, o que fica explícito ao afirmar que “o conceito de materialidade traz um entendimento muito mais rico do caráter público e social da informação em nosso tempo”.

Por entender, assim como Frohmann (2008), que a informação é materizalizada também por meios tecnológicos, é possível afirmar que o FSI pode ser observado e estudado sob esse enfoque. Pode-se buscar a compreensão de como a informação digital, disponibilizada para os voluntários no FSI, é materializada e que tipo de relações e efeitos sociais e políticos essa iniciativa envolve.

Seguindo este pensamento, entende-se que no domínio das pesquisas genealógicas ou de história da família como hoje são mais conhecidas, o estudo do FSI é fundamental, pois segundo Flinders (2012, p. 23) o “a força do programa é sua abordagem colaborativa”. Seguindo esse pensamento, Howe (2009, p. 9) afirma que: “embora o *crowdsourcing* esteja mesclado com a internet, sua essência não é a tecnologia”. Os comportamentos humanos são mais importantes pelo “potencial que a internet tem de interligar a massa da humanidade em um organismo próspero e infinitamente poderoso”.

O domínio do FSI envolve uma ampla comunidade discursiva que pode ser reconhecida, para fins deste estudos, como os que trabalham na digitalização dos registros, os voluntários que indexam os registros, os que atuam como árbitros ao verificar a qualidade das indexações realizadas e todos os que utilizam o *FamilySearch* para a busca de registros e criação de suas árvores familiares.

É inegável que as tecnologias de informação e comunicação ocasionaram inúmeras mudanças na sociedade, mas é importante perceber, como afirma Barreto (2002, p. 73), que “as reais modificações advindas das tecnologias intensas de informação trouxeram ao ambiente um novo elaborar do conhecimento e foram as modificações relacionadas ao tempo e ao espaço de sua passagem”.

4 Considerações finais

A aproximação entre o FSI como iniciativa de *crowdsourcing* e a abordagem sociocognitiva da OC demonstrou a importância dos estudos nesse domínio. Pois, envolve uma comunidade discursiva diversa que, no caso do FSI, é constituída por voluntários sem conhecimento técnico sobre representação descritiva e temática de documentos que se tornarão acessíveis por meio da internet.

É a primeira vez que essa perspectiva é explorada na literatura, portanto, algumas possibilidades de estudo puderam ser identificadas: reconhecer os fatores motivadores para voluntários se envolverem com a indexação no FSI, o que os influencia durante a indexação, quais ferramentas utilizam como apoio para esse processo e de que forma influenciam na OC, qual a função social dessa atividade, quais relações de poder estão envolvidas, entre outras. Além disso, considera-se a análise de domínio uma forma de gerar novo conhecimento sobre a interação dessa comunidade com a informação gerada. Essas podem ser considerações contribuições da OC Para o FSI.

O FSI cresce a cada dia em números de projetos, envolvidos no processo de transcrição e disponibilização de milhões de registros para uma grande comunidade de pesquisa de história da família. Acredita-se que suas principais contribuições para o domínio da OC são: identificação dos relacionamentos dentro dessa comunidade discursiva e como elas influenciam a OC, reconhecimento das influências históricas e sociais entre esses voluntários, elucidação de padrões de busca e uso das informações disponibilizadas e o relacionamento do domínio da OC com um novo domínio, o da história da família ou genealogia.

Referências

ALLEN, JB.; EMBRY, JL.; MEHR, KB. **Hearts turned to the fathers**: a history of the Genealogical Society of Utah, 1894-1994. Provo, Utah: BYU Studies, 1995.

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (org.). **Educação, Universidade e Pesquisa**. Marília: Unesp-Marília-Publicações; São Paulo: FAPESP, 2001. p. 35-60.

BARRETO, A. A.. A condição da informação. **São Paulo em Perspectiva**, v.16, n. 3, 2002. Disponível em: http://www.scielo.br/pdf/spp/v16n3/13563.pdf . Acesso em: 15 out. 2009.

IGREJA DE JESUS CRISTO DOS SANTOS DOS ÚLTIMOS DIAS**. Indexação do FamilySearch,** 2012. Disponível em: https://indexing.familysearch.org/newuser/nuhome.jsf?3.15.1https://familysearch.org/volunteer/indexing Acesso em: 10 set. 2012.

ELLIS, S. A History of collaboration, a future in crowdsourcing: positive impacts of cooperation on british librarianship. **Libri**: International Journal of Libraries and Information Services, v.64, n.1, p.1-10, 2014.

FLINDERS, S. FamilySearch indexing: providing easy access to your records. ARCHIVING, 2012, Copenhagen, Denmark. **Proceedings...** Copenhagen, Denmark: Society for Imaging Science and Technology, jan. 2012. p. 23-27.

FROHMANN, B. O caráter social, material e público da informação. In: FUJITA, M.; MARTELETO, R.; LARA, M. (Orgs). **A dimensão epistemológica da ciência da informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação**. São Paulo: Cultura Acadêmica; Marília: Fundepe, 2008, p.19-34.

HANSEN, D. et al. Improving indexing efficiency & quality: comparing A-B-Arbitrate and peer review. FAMILY HISTORY TECHNOLOGY WORKSHOP, 2012, Salt Lake City, UT. **Proceedings…** Salt Lake City, UT: BYU, 2012. p. 1-4. Disponível em: http://fht.byu.edu/prev_workshops/workshop12/papers/4.1%20DerekHansenEtAlIndexingExperiments.pdf Acesso em: 10 set. 2012.

HANSEN, D. et al. Quality control mechanisms for *crowdsourcing*: peer review, arbitration, & expertise at FamilySearch Indexing. In: CONFERENCE ON COMPUTER SUPPORTED COOPERATIVE WORK, 2013. **Proceedings...** ACM, 2013. p. 649-660. Disponível em: <http://dl.acm.org/citation.cfm?id=2441848> Acesso em: 10 mar. 2015.

HJORLAND, B. Domain analysis in information science. **Journal of Documentation**, Londres, v. 58, n. 4, 2002a, p. 122-162.

HJORLAND, B. Epistemology and the social-cognitive perspective in information science. **Journal of The American Society for Information Science**, v.53, n.4, p. 257-270, 2002b.

HJORLAND, B.; ALBRECHTSEN, H. Toward a New Horizon in Information Science: domain-analysis. **Journal of The American Society for Information Science**, v.46, n.6, p. 400-425. 1995.

HOWE, J. The rise of *crowdsourcing*. **Wired**, n. 1406, p. 1-4, jun. 2006. Disponível em: http://www.wired.com/wired/archive/14.06/crowds.html Acesso em 17 out. 2012.

HOWE, J. **O poder das multidões**. Tradução de Alessandra Mussi Araujo. Rio de Janeiro: Elsevier, 2009.

KENNARD, D.; KENT, A.; BARRETT, W. Linking the past: discovering historical social networks from documents and linking to a genealogical database. In: WORKSHOP ON HISTORICAL DOCUMENT IMAGING AND PROCESSING, 1., 2011, Provo, UT. **ACM International Conference Proceeding Series...** Provo, UT: Brigham Young University, jan. 1, 2011. p. 43-50.

METCALFE, WJ.; THATCHER, MP. Serving the genealogical and historical research communities: an overview of records access and data privacy issues. WORLD LIBRARY AND INFORMATION CONGRESS: IFLA GENERAL CONFERENCE AND COUNCIL, 74., 2008, Quebec, Canadá. **IFLA Conference Proceedings…** Quebec, Canadá: IFLA, 2008. P. 1-10.

TÁLAMO, MFGM.; SMIT, JW. Ciência da Informação: pensamento informacional e integração disciplinar. **Brazilian Journal of Information Science**, v.1, n.1, p. 33-57, jan./jun. 2007. Disponível em: http://www2.marilia.unesp.br/revistas/index.php/bjis/article/viewFile/30/29 Acesso em: 09 out. 2012.

WERSIG, G. Information Science: the study of postmodern knowledge usage. **Information Processing & Management**, v. 29, n. 2, p. 229-239, 1993. http://dx.doi.org/10.1016/0306-4573(93)90006-Y Acesso em: 10 out. 2012.

Estudo de ações para disponibilização de datasets governamentais em linked open data

Fernando de Assis Rodrigues
Universidade Estadual Paulista
fernando@elleth.org

Ricardo César Gonçalves Sant'Ana
Universidade Estadual Paulista
ricardosantana@marilia.unesp.br

1 Introdução

Os princípios de *Linked Open Data* (LOD) estabelecem um novo modo de compartilhamento de *datasets* (conjuntos de dados) abertos pela internet, com objetivo de promover a ampla distribuição de dados estruturados em linguagens como a *eXtensible Markup Language (XML)* e em concomitância com as recomendações do padrão *Resource Description Framework (RDF)* (BERNERS-LEE, 2009; BIZER; HEATH; BERNERS-LEE, 2009; HEATH, 2015; W3C, 2014, 2015).

Neste cenário, *datasets* governamentais têm papel de destaque: representam 18,58% da quantidade total de *datasets* de LOD existentes e 41,54% destes *datasets* governamentais possuem ao menos um vínculo com ontologias ou vocabulários controlados, segundo resultados do mapeamento elaborado pelo projeto *Linking Open Data cloud diagram* (SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Entretanto, segundo Schmachtenberg, Bizer e Paulheim (2014a, p. 12–16) ainda existem características nas estruturas dos *datasets* de LOD no momento da recuperação de dados que não são consideradas ideais ou boa prática adotada, tais como a ausência de metadados e de licenças de uso.

Ações para tornar acessíveis dados governamentais públicos são parte integrante de discussões sobre tendências de modernização dos modelos de administração pública, que procuram redistribuir competências e recursos dentre diferentes organizações

intra e extragovernamentais, permitindo um maior pluralismo institucional nas funções públicas (MALIN, 2006; SANT'ANA; RODRIGUES, 2013).

O fortalecimento de ações de transparência pode ser ampliado com a construção de ambientes de compartilhamento de informação que, dentre outras características, proporcionem a ampliação de fluxos informacionais entre a administração pública e a sociedade, garantindo maior visibilidade das atividades do Estado (BOHMAN, 2000; MARCONDES; JARDIM, 2003). Estes ambientes se tornam componentes de uma maior participação cidadã, estendendo possibilidades de participação além do voto; e o Estado pode melhorar a efetividade e o acompanhamento das atividades e resultados de suas ações, além de atender obrigatoriedade de publicação de dados governamentais (BRASIL, 2011; SANT'ANA; RODRIGUES, 2013).

O acesso aos *datasets* governamentais sobre resultados de votações legislativas é importante no acompanhamento das atividades dos representantes, servindo de apoio a construção de análises, tais como "[…] a identificação de *clusters* partidários" e na "[…] coerência de cada um de nossos representantes nas votações durante seus mandatos" (SANT'ANA; RODRIGUES, 2013, p. 58).

Objetivo e delimitação do tema

O objetivo deste trabalho é o estudo de ações necessárias para a disponibilização de *datasets* governamentais em *Linked Open Data*, a partir da aplicação de um modelo de recomendações para publicação de dados *"Linked Data Best Practices in Different Topical Domains",* proposto por Schmachtenberg, Bizer e Paulheim (2014a), em bases de dados disponíveis sobre votações legislativas do Senado brasileiro.

O objeto de pesquisa foi delimitado a dados disponíveis nas Tecnologias de Comunicação e Informação (TIC) oficiais do Senado brasileiro, mais precisamente sobre votações existentes no 'Portal e-Cidadania – Dados Abertos', analisados entre janeiro e março de 2015.

Procedimentos metodológicos

A metodologia adotada foi a análise exploratória do objeto de pesquisa, de caráter qualitativa, através da explicitação das características: do *dataset* existente (a localização do recurso no web site, informações sobre página descritiva e arquivos disponíveis), e; da estrutura dos dados encontrada no momento da recuperação.
Estas características formaram um conjunto de informações que serviram como subsídio para a proposta de uma estratégia de ações necessárias para a reestruturação destes dados, já existentes, em concomitância com as recomendações estabelecidas de boas práticas de disponibilização de *datasets* em LOD, proposta por Schmachtenberg, Bizer e Paulheim (2014a).

2 Referencial teórico

Schmachtenberg, Bizer e Paulheim (2014a, 2014b) propõem um modelo com recomendações para publicação de dados, com o objetivo de identificar a concomitância aos conceitos de LOD e boas práticas para compartilhamento de dados por *datasets* públicos armazenados em diversos domínios da web. Estas recomendações foram elaboradas a partir das práticas de comunidades e dos resultados apresentados pelo mapeamento de *datasets* LOD elaborado pelo projeto *Linking Open Data cloud diagram* (JENTZSCH; CYGANIAK; BIZER, 2011). O modelo está dividido em nove recomendações:

Fornecimento de identificadores únicos de localização

No processo de recuperação de dados é necessário que *datasets* possuam identificadores únicos para auxiliar o processo de recuperação de dados por agentes externos, em concomitância com o primeiro princípio de LOD (BIZER; HEATH; BERNERS-LEE, 2009). Estes identificadores únicos devem estar em conformidade com regras estabelecidas pelo *Uniform Resource Identifier (URI)* e pelo padrão RDF (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Definição de ligações com outros datasets

É necessário que o *dataset* possua ligações com outros *datasets* através das regras de relacionamentos estabelecidas pelo padrão RDF. Esse procedimento facilita a coleta automatizada por agentes externos de dados disponíveis do *dataset* e dos *datasets* a qual foram vinculados (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Uso de vocabulários controlados e ontologias existentes

Como o dado é um elemento básico "[…] formado por signo ou conjunto finito de signos que não contém, intrinsecamente, um componente semântico, mas somente elementos sintáticos" (SANTOS; SANT'ANA, 2002, s. p.), é necessário o uso de vocabulários controlados e ontologias para ampliar a carga semântica no momento de coleta de dados por agentes externos, tais como o *Dublic Core (DC)*, o *Friend of a Friend (FOAF)*, o *Simple Knowledge Organization System (SKOS)*, entre outros (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Definições de termos, elementos e atributos referenciáveis em vocabulários e ontologias

Os documentos complementares vinculados ao *dataset*, contendo informações sobre ontologias e vocabulários controlados, devem possuir URI únicas para termos, elementos e atributos (SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b). Por exemplo, no FOAF as definições dos termos 'name' e 'birthday' devem ser acessíveis por URI únicas, seja através do uso de símbolos de divisão ('/'), ou uso de *hashtag (#)* para diferenciar o acesso aos termos (BRICKLEY; MILLER, 2014).

Vinculação de termos entre vocabulários

Caso ocorra a necessidade de elaboração de vocabulários próprios, é importante que termos deste novo vocabulário sejam vinculados a termos de vocabulários já existentes e com maior abrangência, como: o DC, o FOAF, o SKOS, entre outros. A vinculação de novos vocabulários com vocabulários abrangentes

proporcionam um maior repertório de informações sobre termos desenvolvidos para o vocabulário próprio aos agentes externos (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Fornecimento de metadados

No momento da recuperação, *datasets* devem ter elementos de metadados para garantir qualidade no processo de recuperação de dados, identificar a origem dos dados e avaliar a qualidade (JENTZSCH; CYGANIAK; BIZER, 2011). Os metadados "[…] são um fator chave para minimizar os problemas de busca e recuperação dos recursos nos diversos ambientes informacionais […]" (SANTOS; ALVES, 2009) e é recomendado que: seus elementos estejam disponíveis no elemento raiz, e; a utilização de elementos do DC (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Uso de licença nos metadados

Os metadados do *dataset* precisam conter licenças de autoria em seus elementos e atributos, tais como: o *Creative Commons, o Open Data Commons Attribution License*, o *Open Database License (ODbL)*, entre outros (JENTZSCH; CYGANIAK; BIZER, 2011; W3C, 2011).

Disponibilização de metadados sobre a estrutura do dataset

Na recuperação do *dataset* devem existir metadados contendo informações sobre a sua estrutura – disponibilizadas junto aos dados ou em documentos suplementares – delimitando elementos, iterações, termos e atributos utilizados (JENTZSCH; CYGANIAK; BIZER, 2011; SCHMACHTENBERG; BIZER; PAULHEIM, 2014a, 2014b).

Uso de métodos alternativos para recuperação de dados

A forma mais comum de disponibilização de *datasets* estruturados no padrão RDF é através de um *SPARQL Endpoint* (JENTZSCH; CYGANIAK; BIZER, 2011), que habilita agentes externos a realizar buscas estruturadas na linguagem de consulta *SPARQL*. Todavia, é recomendado que também estejam disponíveis arquivos de despejo (*RDF Dump*), explicitados no padrão RDF/XML ou equivalente (SEMANTICWEB.ORG, 2011).

Características do *dataset*

O Portal e-Cidadania tem o objetivo de promover transparência de ações e atividades do Senado Brasileiro, através do acesso a dados (BRASIL, 2015a). Em janeiro de 2015, o *web site* possuía quarenta e cinco *datasets*, divididos em oito grupos: 'Projetos e Matérias', 'Plenário', 'Parlamentares', 'Composição', 'Comissões', 'LexML', 'Legislação', e 'Processo Legislativo'.

O grupo 'Plenário' contém oito subdivisões: 'Diários do Senado' e do Congresso'; 'Legislaturas e 'Sessões Legislativas'; 'Matérias com prazos'; 'Pronunciamentos de senador'; 'Questões de Ordem'; 'Sessões do Plenário'; 'Tabelas de tipos relacionados a plenário', e; 'Votações nominais' – este último contendo dados sobre "[…] votos registrados nas sessões de votação em plenário […]" e "[…] informações sobre as sessões, as matérias votadas, totalização de votos e o voto de cada parlamentar" (BRASIL, 2015b).

Votações nominais é composto por 11 itens: 9 arquivos de despejo no formato XML, contendo dados sobre votações agrupadas anualmente; 1 hiperlink para um *web service* de consulta automatizada, e; 1 hiperlink para uma página de consulta de votações – este último, por se tratar de consultas em formato *HyperText Markup Language (HTML)* não é objeto deste estudo.

Características da estrutura dos dados na recuperação

Cada arquivo de despejo no formato XML possui uma URL para identificação, formado pela composição: domínio/hierarquia '*http://legis.senado.leg.br/dadosabertos/dados/'; prefixo 'ListaVotacoes*', ano referente aos dados, e; extensão de arquivo *'.xml'*.

Na recuperação de dados via *web service*, os *datasets* apresentam apenas resultados agrupados por consultas diárias. Por exemplo, para realizar uma coleta de dados sobre votações ocorridas em um determinado mês, é necessário realizar ‘x’ consultas, onde ‘x’ representa o total de dias do mês. Essa recuperação de datasets possui uma URL para identificação de resultado diário, cujo valor da URL é formado pela composição: domínio/hierarquia ‘*http://legis.senado.leg.br/dadosabertos/plenario/lista/votacao/";* ano; mês, e; dia.

Em ambos os casos – na coleta dos arquivos de despejo e na recuperação de dados via *web service* – os *datasets* estão explicitados na linguagem XML e as coleções de elementos, atributos e termos disponíveis são idênticas (Quadro 1).

Elemento	Associado ao elemento	Tipo de Dado	Atributos
ListaVotacoes	Nenhum (Elemento Raiz)	Elem. Agrupador	'xmlns:xsi' e 'xsi:noNamespaceSchemaLocation'
Metadados	ListaVotacoes	Elem. Agrupador	Nenhum
Votacoes	ListaVotacoes	Elem. Agrupador	Nenhum
Versao	Metadados	Texto	Nenhum
VersaoServico	Metadados	Número inteiro	Nenhum
DescricaoDataSet	Metadados	Texto	Nenhum
Votacao	Votacoes	Elem. Agrupador	Nenhum
CodigoSessao	Votacao	Número inteiro	Nenhum
SiglaCasa	Votacao	Texto	Nenhum
CodigoSessaoLegislativa	Votacao	Número inteiro	Nenhum
TipoSessao	Votacao	Texto	Nenhum
NumeroSessao	Votacao	Número inteiro	Nenhum
DataSessao	Votacao	Texto	Nenhum
HoraInicio	Votacao	Texto	Nenhum
CodigoTramitacao	Votacao	Número inteiro	Nenhum
CodigoSessaoVotacao	Votacao	Número inteiro	Nenhum
SequencialSessao	Votacao	Número inteiro	Nenhum
Secreta	Votacao	Texto	Nenhum
DescricaoVotacao	Votacao	Texto	Nenhum
Resultado	Votacao	Texto	Nenhum
TotalVotosSim	Votacao	Número inteiro	Nenhum
TotalVotosNao	Votacao	Número inteiro	Nenhum
TotalVotosAbstencao	Votacao	Número inteiro	Nenhum
CodigoMateria	Votacao	Número inteiro	Nenhum
SiglaMateria	Votacao	Texto	Nenhum
NumeroMateria	Votacao	Número inteiro	Nenhum
AnoMateria	Votacao	Número inteiro	Nenhum
Votos	Votacao	Elem. Agrupador	Nenhum

VotoParlamentar	Votos	Elem. Agrupador	Nenhum
CodigoParlamentar	VotoParlamentar	Número inteiro	Nenhum
NomeParlamentar	VotoParlamentar	Texto	Nenhum
SexoParlamentar	VotoParlamentar	Texto	Nenhum
Url	VotoParlamentar	Texto	Nenhum
Foto	VotoParlamentar	Texto	Nenhum
Tratamento	VotoParlamentar	Texto	Nenhum
Voto	VotoParlamentar	Texto	Nenhum

Quadro 1 – Estrutura de elementos e atributos do *dataset*

Fonte: Elaborado pelos autores

O elemento raiz é denominado ‘ListaVotacoes’ e possui dois atributos e dois elementos. Seus dois atributos são responsáveis por vincular o *dataset* com um documento suplementar (*XML Schema*), contendo a delimitação de elementos, tipos de conteúdos e atributos disponíveis.

Seus elementos a) ‘Metadados’ e b) ‘Votacoes’ são elementos agrupadores com seu valor sendo um conjunto de um ou mais elementos; ambos sem atributos.
O elemento agrupador ‘Metadados’ possui três elementos; ‘Votacoes’ contém um ou mais elementos ‘Votacao’; e o elemento ‘Votacao’ possui dezenove elementos. Nenhum dos elementos possui atributos.

O elemento agrupador ‘Votos’ (filiado ao elemento ‘Votacao’) contém um ou mais elementos ‘VotoParlamentar’, sem atributos. O elemento ‘VotoParlamentar’ possui sete elementos, também sem atributos.

3 Resultados

A partir da análise, foram propostas oito ações para a disponibilização de *datasets* governamentais em LOD a partir da recuperação de dados já existente (Figura 1).

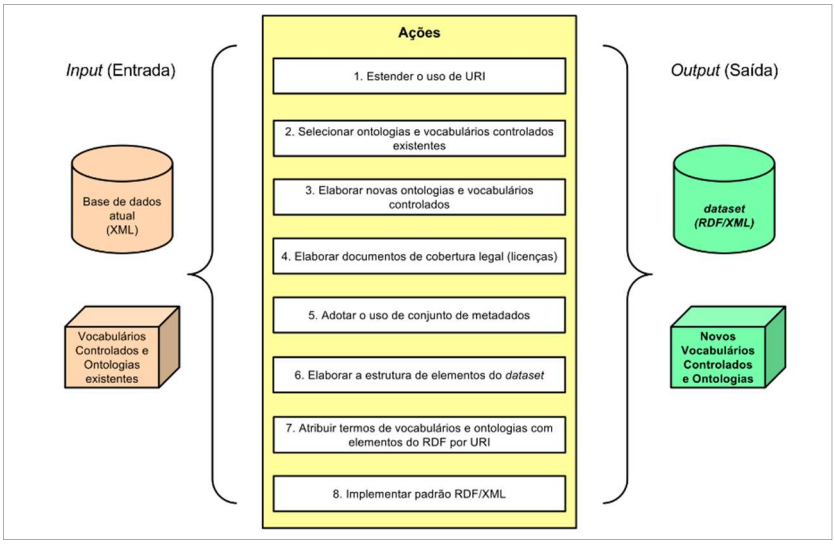

Ações
Input (Entrada)
Output (Saída)
1. Estender o uso de URI
2. Selecionar ontologias e vocabulários controlados existentes
3. Elaborar novas ontologias e vocabulários controlados
4. Elaborar documentos de cobertura legal (licenças)
5. Adotar o uso de conjunto de metadados
6. Elaborar a estrutura de elementos do dataset
7. Atribuir termos de vocabulários e ontologias com elementos do RDF por URI
8. Implementar padrão RDF/XML
Base de dados atual (XML)
Vocabulários Controlados e Ontologias existentes
dataset (RDF/XML)
Novos Vocabulários Controlados e Ontologias

Figura 1 – Síntese de ações necessárias ao desenvolvimento de *datasets* governamentais

Fonte: Elaborado pelos autores.

Anteriormente a execução destas ações é importante que estejam disponíveis informações *(input)* sobre as características das bases de dados já existentes e um conhecimento prévio dos envolvidos sobre ontologias e vocabulários existentes que possam ser parte da explicitação de relacionamentos e de características de elementos do novo *dataset* LOD.

As ações identificadas neste trabalho podem ser resumidas em:

- Ação 1: estender o uso de identificadores URI para a identificação de elementos, atributos e termos que compõem o RDF, explicitados em documentos suplementares ao *dataset* para a compreensão das regras deste sistema;
- Ação 2: selecionar ontologias e vocabulários controlados de uso por muitas comunidades que possam ser parte da explicitação de relacionamentos e elementos do novo *dataset* LOD;
- Ação 3: elaborar ontologias e vocabulários específicos para a explicitação de relacionamentos e elementos não existentes em ontologias e vocabulários da segunda ação;
- Ação 4: elaborar documentos complementares contendo a cobertura legal como licenças de uso e direitos autorais, e vincular estes documentos com o *dataset.* É importante que estas licenças sejam explicitadas nos metadados (quinta ação);

- Ação 5: adotar o uso de conjuntos de elementos de metadados de iniciativas populares para explicitar o conteúdo do *dataset* na recuperação de dados por agentes externos;
- Ação 6: elaborar a estrutura lógica dos elementos do *dataset*, atributos, títulos de elementos, valores, agrupadores e regras de validação;
- Ação 7: atribuir termos das ontologias e vocabulários selecionados aos elementos do *dataset*, vinculando-os através de URI, para ampliar a carga semântica destes dados no momento de coleta;
- Ação 8: implementar as regras do padrão RDF sobre a estrutura da linguagem de marcação XML, respeitando as formas estabelecidas para documentos RDF/XML.

Estas ações devem desenvolver como resultado *(output)* o *dataset* LOD (explicitado em formato RDF/XML ou equivalente); e vocabulários e ontologias elaboradas para o atendimento das necessidades do contexto dos dados (votações nominais).

4 Conclusões

A forma de recuperação atual dos conjuntos de dados não contempla características importantes na contextualização dos dados, como o uso de vocabulários controlados e de ontologias, interferindo diretamente na independência do agente externo na coleta dos dados.

Apesar da existência de três elementos de metadados não existem informações como: autor, licença, fonte, data de criação e data de publicação.

A aplicação do modelo de recomendação serviu como norteador para a formação de ações propostas neste estudo, principalmente ao fornecer subsídios para elaboração do conjunto de ações necessárias para disponibilização de *datasets* governamentais em LOD*,* fornecendo aos gestores públicos pontos de intervenção necessários para que dados já disponíveis seus *websites* possam ser reestruturados em *datasets* LOD e, consequentemente, passíveis de coleta por agentes externos.

Espera-se que a aplicação destas ações em dados sobre votações sirva de estímulo para a aplicação destas ações em outras bases de dados, em outras esferas

e em outros *websites*; como também estimular o surgimento de novos *datasets* governamentais brasileiros nesta temática.

Referências

BERNERS-LEE, T. **The next Web**. In: TED2009. Estados Unidos da América: fev. 2009.

BIZER, C.; HEATH, T.; BERNERS-LEE, T. Linked Data – The Story So Far: **International Journal on Semantic Web and Information Systems**, v. 5, n. 3, p. 1–22, 33 2009.

BOHMAN, J. **Public deliberation: Pluralism, complexity, and democracy**. [s.l.] MIT press, 2000.

BRASIL. 12.527. Lei Nº 12.527, de 18 de Novembro de 2011 (Lei de Acesso à Informação). . 18 nov. 2011.

BRASIL. **Portal e-Cidadania**, 2015a. Disponível em: <http://dadosabertos.senado.gov.br>. Acesso em: 3 abr. 2015

BRASIL. **Votações Nominais**, 2015b. Disponível em: <http://dadosabertos.senado.gov.br/dataset/votaces-nominais>. Acesso em: 3 abr. 2015

BRICKLEY, D.; MILLER, L. **FOAF Vocabulary Specification 0.99**, 14 jan. 2014. Disponível em: <http://xmlns.com/foaf/spec/20140114.html>. Acesso em: 27 mar. 2015.

HEATH, T. **Frequently Asked Questions**Linked Data - Connect Distributed Data across the Web, , 2015. Disponível em: <http://linkeddata.org/faq>. Acesso em: 22 mar. 2015

JENTZSCH, A.; CYGANIAK, R.; BIZER, C. **State of the LOD Cloud**, set. 2011. Disponível em: <http://lod-cloud.net/state/>. Acesso em: 24 mar. 2015

MALIN, A. M. B. Gestão da Informação Governamental: em direção a uma metodologia de avaliação. **DataGramaZero - Revista de Ciência da Informação**, v. 7, n. 5, out. 2006.

MARCONDES, C. H.; JARDIM, J. M. Políticas de Informação Governamental: a construção de Governo Eletrônico na Administração Federal do Brasil. **DataGramaZero - Revista de Ciência da Informação**, v. 4, n. 2, abr. 2003.

SANT'ANA, R. C. G.; RODRIGUES, F. DE A. Visualização de afinidades entre parlamentares mediante dados de votações no Senado Brasileiro. **Informação & Sociedade: estudos**, v. 23, n. 1, p. 49–59, jan. 2013.

SANTOS, P. L. V. A. DA C.; ALVES, R. C. V. Metadados e Web Semântica para

estruturação da Web 2.0 e Web 3.0. **DataGramaZero – Revista de Ciência da Informação**, v. 10, n. 6, dez. 2009.

SANTOS, P. L. V. A. DA C.; SANT'ANA, R. C. G. Transferência da Informação: análise para valoração de unidades de conhecimento. **DataGramaZero - Revista de Ciência da Informação**, v. 3, n. 2, abr. 2002.

SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. Adoption of the Linked Data Best Practices in Different Topical Domains. In: MIKA, P. et al. (Eds.). . **The Semantic Web – ISWC 2014**. Cham: Springer International Publishing, 2014a. v. 8796p. 245–260.

SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. **State of the LOD Cloud 2014**University of Mannheim, 30 ago. 2014b. Disponível em: <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>. Acesso em: 25 mar. 2015

SEMANTICWEB.ORG. **SPARQL endpoint**Semantic Web.org, , 26 nov. 2011. Disponível em: <http://semanticweb.org/index.php?title=SPARQL_endpoint&oldid=52541>. Acesso em: 4 jan. 2015

W3C. **Data Licensing**W3C, 11 abr. 2011. Disponível em: <http://www.w3.org/wiki/index.php?title=TaskForces/CommunityProjects/LinkingOpenData/DataLicensing&oldid=49411>. Acesso em: 20 mar. 2015

W3C. **RDF 1.1 XML Syntax**W3C, 25 fev. 2014. Disponível em: <http://www.w3.org/TR/2014/REC-rdf-syntax-grammar-20140225/>. Acesso em: 22 mar. 2015

W3C. **Linked Data Platform 1.0**W3C, 26 fev. 2015. Disponível em: <http://www.w3.org/TR/2015/REC-ldp-20150226/>. Acesso em: 24 mar. 2015

Organização do conhecimento e o patrimônio industrial em São Paulo: o Projeto Eletromemória

Vânia Mara Alves Lima
Universidade de São Paulo
vamal@usp.br

Marcia Cristina de Carvalho Pazin Vitoriano
Universidade Estadual Paulista
marciapazin@marilia.unesp.br

Cristina Hilsdorf Barbanti
Universidade de São Paulo
crishb@usp.br

1 Introdução

Este trabalho discute e analisa as questões referentes à organização do conhecimento sobre o patrimônio industrial, no domínio específico História da Energia Elétrica no Estado de São Paulo, com objetivo de representar, recuperar e disseminar a informação que se encontra em acervos documentais dispersos, a partir de sua disponibilização, em um instrumento de pesquisa em ambiente web.

Essas questões surgiram no desenvolvimento de projeto temático, financiado pela FAPESP em duas fases, voltado ao levantamento do patrimônio histórico do setor de geração, transmissão e distribuição de energia elétrica no estado de São Paulo ao longo do século XX. Na primeira fase, denominada Eletromemória I, de 2007 a 2011 foram realizados levantamentos e análises das unidades de geração, transmissão e

distribuição de energia elétrica, construídas durante as décadas de 1950-1970, representativas do período de consolidação do sistema elétrico estatal. Na segunda fase do projeto, denominada Eletromemória II, que se iniciou em 2012, vem sendo estudadas cerca de 50 unidades de geração de energia elétrica, entre usinas e pequenas centrais hidrelétricas (PCHs), construídas no período de 1890 a 1960. Interdisciplinar e interinstitucional a equipe do projeto conta com professores, pesquisadores e estudantes das três universidades paulistas – USP, UNESP e Unicamp das áreas de História, Geografia, Museologia, Arquivologia, Biblioteconomia e Arquitetura.

Considerando as diversas questões envolvidas na implantação de empreendimentos dessa natureza, o Eletromemória II ao contemplar aspectos históricos, documentais, ambientais e de patrimônio industrial, definiu quatro eixos temáticos de atuação, onde cada um deles aborda questões relativas à área do conhecimento a que se refere, a saber:

- **História**: volta-se à pesquisa sobre as transformações das unidades produtivas e de seus equipamentos, considerando a história da tecnologia, além da história dos trabalhadores que vivenciaram a operação dessas usinas em diferentes momentos.
- **Patrimônio Industrial/Cultura Material/Museologia**: volta-se para a pesquisa das condições do patrimônio representado por edificações, equipamentos e demais elementos da cultura material identificados nas unidades estudadas, numa perspectiva de estudo dos remanescentes do patrimônio industrial do estado de São Paulo e de seu potencial museológico.
- **Paisagem/Meio-Ambiente**: baseado numa abordagem descritivo-comparativa, volta-se ao estudo da paisagem do entorno dos locais de implantação das unidades estudadas, considerando a fauna e a flora originais e suas condições atuais, as alterações do regime de águas e o histórico de uso e ocupação das áreas.
- **Ciência da Informação**: volta-se à identificação dos registros documentais relativos às usinas estudadas e à definição de parâmetros

> para a representação da informação relativa ao conhecimento acumulado sobre cada uma das unidades analisadas.

Ao agregar diferentes áreas do conhecimento, num exemplo de comunidade de pesquisa interdisciplinar, o projeto necessita de um instrumento de busca que consolide os dados obtidos. Neste trabalho abordamos a elaboração do Inventário de Patrimônio Industrial da Energia Elétrica no Estado de São Paulo, instrumento que consolida a pesquisa sob dois aspectos: a uniformização terminológica e a uniformização da estrutura descritiva de conteúdos.

Como hipótese de trabalho, consideramos que a mediação entre sistemas de organização do conhecimento em comunidades distintas pode acontecer no nível das Linguagens Documentárias e da Terminologia. Nesse sentido, é necessária a elaboração de ferramentas que articulem e transponham domínios de conhecimento e promovam pontes reais de comunicação entre seus diversos agentes.

O referencial teórico-metodológico utilizado encontra-se nas fronteiras epistemológicas da Arquivologia, Biblioteconomia e Museologia reunidas sob a ótica da Ciência da Informação. Trata-se de um relato de experiência, de uma pesquisa exploratória de abordagem qualitativa.

2 A padronização terminológica

O trabalho do grupo do eixo Ciência da Informação, durante o Eletromemória I teve como objetivo a elaboração de um vocabulário controlado para representar e recuperar a informação armazenada em acervos documentais pertencentes especificamente ao domínio da energia elétrica. A premissa básica era a de que o controle terminológico garantiria a representação conceitual do domínio, ainda que a interpretação fosse submetida a pontos de vista de comunidades diversas e condicionada às características específicas da ampla tipologia documental do universo focalizado. O vocabulário controlado foi elaborado a partir dos catálogos de assuntos dos acervos arquivístico, bibliográfico e museológico da Fundação Energia e Saneamento e teve sua primeira versão incluída na base de dados Enerweb que se encontra disponibilizada na página web http://www.energiaesaneamento.org.br/.

O Eletromemória II amplia o mapeamento dos acervos documentais, e insere novos eixos de pesquisa, tendo como objetivo atualizar o instrumento de controle terminológico, incluindo termos referentes ao patrimônio industrial do setor elétrico que serão utilizados para descrição das unidades produtivas de característica histórica. Sabemos que, principalmente a partir da década de 1990, no ambiente empresarial de concessões, a própria regulação da atividade econômica trouxe uma padronização mínima das referências terminológicas dos acervos. A existência desses padrões documentais mínimos para atendimento à legislação fiscalizatória possibilitou a interlocução entre acervos de distintas empresas do ponto de vista da terminologia utilizada. Porém, a ocorrência desse tipo de controle para os acervos do início da implantação da energia elétrica, entre o final do século XIX e meados do XX, ainda é extremamente limitada, não existindo o controle de terminologia para a produção documental ou para outras questões técnicas do patrimônio, como equipamentos e técnicas construtivas.

Uma das características dos instrumentos para controle de vocabulário é, também, a possibilidade de integração terminológica em diferentes áreas do conhecimento. A associação da pesquisa nos dois eixos – informação e documentos – possibilitará a criação de um instrumento mais abrangente, tanto do ponto de vista dos conteúdos informados quando do controle terminológico da informação, possibilitando ao pesquisador acesso amplo e qualificado ao acervo.

As pesquisas de campo que vem sendo realizadas nas unidades geradoras de eletricidade, e em acervos públicos e privados, com objetivo de diagnosticar o patrimônio arquivístico, industrial, de cultura material e de potencial museológico, indicando seu estado de organização e conservação, além de sua relação geograficamente com o entorno e sua transformação no tempo, permitem a coleta de novos termos a serem inseridos no vocabulário controlado.

Para proceder a inclusão desses novos termos realizou-se um encontro entre os pesquisadores do projeto para a elaboração de um mapa conceitual, conforme definido por Fendrich e Pereira (2006) e Rovira (2005), isto é, foram organizadas proposições sobre o tema principal do projeto (História da Energia Elétrica) que, visualizado graficamente permitiu o entendimento das relações entre os conceitos conectados estimulando a solução de problemas e possibilitando a percepção de diferentes caminhos dentro deste sistema conceitual.

As discussões possibilitaram identificar os termos: **Patrimônio industrial; Paisagem e Meio-ambiente**; **Processos de eletrificação** e **Documentação** como subcategorias da categoria **História da Energia Elétrica** a ser reestruturada no vocabulário controlado. Todas as subcategorias devem ser organizadas a partir do estabelecimento das relações lógico-semânticas entre os termos que a compõem. Essas relações são estabelecidas a partir das definições de cada termo identificado como pertencente a essa subcategoria. A definição de um termo deve ser elaborada a partir do que é prescrito pelas normas terminológicas ISO 704 (2000) e a 1087 (2000).

Segundo Dahlberg (1978), a definição de um termo deve enunciar as características essenciais do conceito, isto é, as características sem as quais não conseguimos compreender o conceito que o termo designa. A definição deve também enunciar as características comuns, ou seja, aquelas que permitem que dois termos estejam na mesma subcategoria/categoria, assim como as características disjuntivas, aquelas que estabelecem os limites dessa subcategoria/categoria.

A utilização da terminologia de domínio como referência para um vocabulário controlado garante o controle da significação de seus termos e consequentemente garante a qualidade da informação por ele representada. Para garantir a interoperabilidade futura com outros instrumentos de controle terminológico utiliza-se a norma ISO 25964-1 (2011) para elaboração de tesauros e interoperabilidade entre vocabulários.

Assim, espera-se que os termos definidos e relacionados adequadamente, garantam o significado dos metadados que constituem o instrumento de pesquisa Inventário do Patrimônio Industrial de Energia Elétrica de São Paulo e ao mesmo tempo garantam o conteúdo dos documentos ali descritos.

3 O Inventário de Patrimônio Industrial de Energia Elétrica em São Paulo

Trata-se de um instrumento de pesquisa que articula o conhecimento produzido por áreas distintas, referindo-se a um mesmo objeto de pesquisa. Após a análise dos conceitos de cada área, ficou claro para os pesquisadores que seria necessário articular os conteúdos de modo a desenvolver uma estrutura de dados comum que atendesse a critérios mínimos de representação da informação de cada um dos eixos

temáticos tratados, preservando suas especificidades. Por isso, a definição da estrutura de metadados deveria considerar a existência de:

1. Necessidades informacionais relativas a cada área do conhecimento;
2. Padrões normativos que atendessem a essas necessidades e que possibilitassem a articulação dos conteúdos em dois níveis: informações de diferentes naturezas sobre uma mesma unidade e informações da mesma natureza sobre diferentes unidades.

A análise comparativa realizada levou em conta questões como a grande abrangência geográfica e temporal de um projeto dessa natureza. Considerando o universo de pesquisa, os pesquisadores deveriam compatibilizar informações sobre unidades espalhadas por todas as regiões do estado de São Paulo, construídas em momentos diferentes, tendo, portanto, características específicas, fossem elas construtivas, funcionais ou ambientais. Além disso, questões de natureza administrativa, como a transição da propriedade entre diversas empresas e órgãos governamentais também deveriam ser consideradas na construção do instrumento.

Após analisar as possibilidades existentes nas áreas de Arquivologia, Museologia e Biblioteconomia percebeu-se que, para a finalidade do projeto, seria interessante utilizar um instrumento descritivo que pudesse ser aplicado a qualquer uma das unidades, descrevendo-a em diversos níveis, o que corresponde ao conceito de descrição multinível utilizado pela norma internacional de descrição arquivística, a ISAD-G (CONSELHO INTERNACIONAL DE ARQUIVOS, 2000).

Apesar de tratar-se de um instrumento de pesquisa mais abrangente do que um guia de acervo tradicional, definido como um “instrumento de descrição documental que permite uma visão de conjunto de um serviço de arquivo ou de um organismo e que apresenta informações práticas sobre o conteúdo e a organização de cada um dos fundos”, (CUNHA; CAVALCANTI, 2008, p.183), ou numa acepção da biblioteconomia, “um documento com instruções para orientar os usuários, sobre o conhecimento e exploração do acervo de organismos documentários” (CUNHA; CAVALCANTI, 2008, p.183), percebemos que o conceito poderia ser aplicado a este caso, sendo obviamente necessário um alargamento dos conceitos a serem aplicados. A ideia de que cada unidade produtiva poderia ser considerada como uma unidade relativamente independente, configurando-se como um conjunto individualizado, tornou esta opção a mais interessante. Desta forma, é possível tratar

todo o acervo descrito de um ponto de vista documental, especialmente considerando-se o contexto de pesquisa em que o projeto se insere.

Ao mesmo tempo, o instrumento final teria que consolidar o conceito de inventário utilizado pela área de Patrimônio Industrial, de forma a torná-lo reconhecível para todos os pesquisadores. O “registro dos bens e demais coisas pertencentes a uma pessoa ou comunidade” (RETTIG, 2009, p. 190) deve seguir as orientações da Carta de Nizhny Tagil, que indica a necessidade de levantamentos de campo, a elaboração de tipologias industriais e de inventários dos sítios, que proporcionem acesso fácil e livre ao público de interesse. Também orienta a inclusão de descrições, desenhos e fotografias, além da indicação de fontes documentais (TICCH, 2003).

Visando atender as especificidades apresentadas anteriormente, o instrumento foi dividido em seis grandes áreas informacionais:

1. Fundo / Coleção - Identificação da unidade produtiva: contendo dados básicos de identificação e propriedade, dados construtivos, localização geográfica, datas de início e fim de operação. Essa área compatibiliza informações de todos os campos específicos do conhecimento, considerando a acumulação e os direitos sobre o acervo. O termo Fundo, utilizado na Arquivologia, remete a questões similares ao termo Coleção, utilizado na Biblioteconomia e Museologia.

   Dentro deste item, incorporamos questões como a autoria, a proveniência e a procedência dos acervos, considerando questões teóricas da Arquivologia, mas que foram adaptados para atender às características metodológicas das diferentes áreas.
2. Histórico da unidade: com informações sobre a história da unidade produtiva, desde as motivações para sua criação, importância para a região e dados cronológicos sobre a unidade produtiva;
3. Patrimônio industrial / Museologia: com a descrição da estrutura física remanescente (tanto arquitetônica, quando de equipamentos) e aspectos museológicos, como a análise dapossibilidade de implantação de um roteiro de visitação;
4. Paisagem / Meio Ambiente: com análise comparativa das características ambientais do entorno da unidade estudada;

5. Acervo arquivístico: contendo as informações sobre a localização de documentos de valor histórico, suas condições de preservação e acesso.
6. Palavras-chave: todos os itens de pesquisa passaram por análise e padronização terminológica, constituindo descritores, formatados a partir do vocabulário controlado definido. Dessa forma, a normalização terminológica é a responsável por compatibilizar os conteúdos entre as diversas unidades pesquisadas, possibilitando a análise comparativa em diferentes contextos.

3 Conclusões

Considerando a hipótese de trabalho desta pesquisa, onde a terminologia e a estrutura informacional são responsáveis pela mediação de sistemas de organização de conhecimento em comunidades distintas, o trabalho apresentou um relato de experiência em que o desenvolvimento do instrumento de representação da informação incorpora e articula conceitos de diversas áreas do conhecimento.

A análise comparativa de conceitos utilizados pelas diferentes áreas sobre o tema História da Energia Elétrica no Estado de São Paulo e a organização das referências informacionais sobre o tema possibilitaram a criação de um instrumento descritivo em que se aliam o controle terminológico e o registro dos conteúdos.

Por outro lado, a elaboração de um instrumento único, apresentando informações específicas de cada área do conhecimento sobre um mesmo objeto, articuladas entre si, possibilita ao pesquisador uma visão abrangente e interdisciplinar da unidade pesquisada, no caso, cada pequena central hidrelétrica participante do projeto.

Entendemos que, com isso, tanto o conhecimento sobre cada unidade, quanto as relações estabelecidas entre elas, tornam-se mais claros e passíveis de utilização consistente pelas diversas comunidades de pesquisadores envolvidas.

Referências

CONSELHO INTERNACIONAL DE ARQUIVOS. ISAD(G): **Norma geral internacional de descrição arquivística**: segunda edição, adotada pelo Comitê de Normas de Descrição, Estocolmo, Suécia, 19-22 de setembro de 1999, versão final aprovada pelo CIA. Rio de Janeiro: Arquivo Nacional, 2000.

CUNHA, M.B; CAVALCANTI, C.R.O. **Dicionário de Biblioteconomia e arquivologia. Brasília**, DF: Briquet de Lemos / Livros, 2008.

DAHLBERG, I. Teoria do conceito. **Ciência da informação**. Rio de Janeiro, v.7, n.2, p.101-107, 1978.

FENDRICH, L J; PEREIRA, L**.** Ensinar e Aprender no Ensino Superior Através de Mapas Conceituais In**:** SIMPEP, 8., 2006. **Anais**, Bauru.

INTERNATIONAL STANDARD ORGANIZATION.**ISO 1087**: **Terminology work - vocabulary**. . Geneva : ISO, 2000

INTERNATIONAL STANDARD ORGANIZATION.**ISO 25964-1**: **Information and documentation - Thesauri and interoperability with other vocabularies – part 1** – Thesauri for information retrieval. Geneva : ISO, 2011.

INTERNATIONAL STANDARD ORGANIZATION.**ISO 704**: **terminology work - principles and methods**. Geneva : ISO, 2000

RETTIG, J.M. **Glossário Sextilingüe de Patrimonio Industrial**. Santiago de Chile: Universidad Internacional SEK / TICCIH-Chile, 2009.

THE INTERNATIONAL COMMITTEE FOR THE CONSERVATION OF THE INDUSTRIAL HERITAGE (TICCIH).**Carta de NizhnyTagilsobre o patrimônio industrial**.TICCIH, 2003. Disponível em: http://ticcih.org/wp-content/uploads/2013/04/NTagilPortuguese.pdf Acesso em 30/04/2015.

A Dimensão Social, Cultural e Política da Organização e Representação do Conhecimento

Cultural Interoperability and Knowledge Organization Systems

Widad Mustafa El Hadi
University of Lille 3
widad.mustafa@univ-lille3.fr

1. Context and Rationale

This article addresses cultural interoperability in a context of rationalization. That means, it aims at the same time at taking a certain distance from rationalization questions and the challenges with which KO community is confronted. Examining cultural interoperability would involves other questions: how would it be possible to avoid falling into a downward technicization spiral since interoperability, especially the technical one, is rooted in the classic computing problems underpinning information systems compatibility from the point of view of their formats, data, interchange protocols and so on. What hinders semantic and consequently cultural interoperability is that the degree of success that can be achieved in the integration of multiple knowledge representation systems or knowledge organization schemes is constrained by limitations on the universality of human conceptual systems. For example, human languages do not all lexicalize the same set of concepts; nor do they structure (quasi-)equivalent concepts in the same relational patterns. As a consequence, even multilingual thesauri designed from the outset from the perspective of multiple languages may routinely include situations where corresponding terms are not truly equivalent (Hudon, 1997), 2005, 2012), (Green & al. 2002).

It is therefore essential to pay attention to the local, organizational, professional or national cultures, in KOS tools design and implementation, user studies, semantic Web technology and "intercultural" interfaces design. One of the questions we are addressing is, from where do culture of cultural interoperability grow? It will be interesting to examine intercultural exchange contexts: from technical interoperability

to intercultural interoperability through semantic, military and enterprise interoperability.

2. Outline

The core issue in my presentation is Cultural interoperability in knowledge organization systems (KOS). Before examining this question I will first revisit the concept of interoperability and its layers with a special focus on semantic interoperability. This investigation is partly based on a previous paper (Favier & Mustafa El Hadi, 2013) where we have considered that semantic interoperability originated on one hand from KOS Translatability, and on the other hand from navigating within their diversity through interfaces linking one or two KOS.[4] This will lead us to bring forth the following hypothesis: that semantic interoperability has a broad scope or coverage and that this very semantic interoperability is one of the core elements towards cultural interoperability. It should be noted however, that full semantic interoperability cannot operate in all contexts in spite of the efforts made in that direction. As a result there is even more to say about the difficulties ahead in trying to achieve “cultural interoperability”. The main question I will try to address here, is to what extent can this concept apply to KOS?

The article will be structured as follows: I will first review the ongoing research on KOS interoperability and define the concept of interoperability in its many layers (Miller 2002). I will then define “cultural interoperability” in its different aspects: within the military sphere, in diplomacy or in enterprise activities before trying to circumscribe it to KOS.

The term knowledge organization system (KOS) covers all types of schemes for organizing information and promoting knowledge management. They include classification schemes that organize materials at a general level (such as books on a shelf), subject headings providing more detailed access, and authority files which control different versions of key information (geographical names or personal names).

[4] Our experience is based on the evaluation of the Visual Catalog (Papy, 2013) which links classification data using DDC, to RAMEAU authority file and other essential UNIMARC data. It is currently used by six French universities. The Visual Catalog is not based on translability or semantic features but on an interface mapping between numerous vocabularies tailored for search and retrieval in education knowledge-domain. Two sub-domains vocabularies pertaining to schools and are enhancing mapping. We considered these two sub-domains as representing to two discourse communities or "cultures" the Visual Catalog is thus used as a hub linking the two vocabularies.

This also includes less traditional schemes, such as semantic networks and ontologies (Hodge 2000). Because knowledge organization systems are mechanisms for organizing information, they are at the core of every library, museum, and archive. One of the challenges we face in knowledge organization and representation is heterogeneity of systems both at the level of expressions and structure of conceptual content (see Green 2002, Doerr 2001; Chun & Moen 2001; Zeng & al. 2004, among many others). This point will be extensively developed in section 5.

3. The Importance of KOS Interoperability in Digital Environments

The pervasive power of digitization causes scientific, educational, economic and cultural communities to change their modes of accessing, sharing and disseminating knowledge. As a result this leads to a convergence between our cultural heritage, classic culture and technical culture. It is no surprise that some researchers have called for a "*digital humanism*" (Doueihi 2011), pointing out the way in which new technologies are becoming a sort of «culture» which "drives us into a new global cultural destiny" (Mustafa El Hadi, 2013). This concerns society as a whole. One of the driving forces of these developments is the scientific community i.e. academia which is both "maker" and "consumer" of these technologies.

General search engines on the Web are being so successful in performing the tasks they are designed for that they create an illusion of ease and efficiency that gradually shapes public opinion and influences decisions having to do with investment into information services, education, research or culture (DIGIKO, 2013).

Yet, the most important characteristics of the information environment speak of an increasing complexity of information discovery:

- information on the Web are shared globally (multilinguality and multiculturality)

- Web is participatory in nature: experts and non-experts are involved in information production, information management and access

-information seeking scenarios are becoming multiple and unpredictable, with users who have different requirements

- the information sought is now heterogeneous, comprising text, images, sounds or datasets

- hybrid collections now contain both digital and non digital resources , with full text resources and metadata[5] pointing to resources

To date, most of the efforts in digital environment have been focused on technical interoperability issues: architecture, metadata standards, copyright and interoperability that are crucial to the management of digital repositories. Thus, the bulk of interoperability research has focused on technical and informational issues. Until recently, relatively little effort or expertise has been directed towards matters of semantic content access, semantic interoperability, or knowledge capture and facilitation. Even less attention is thus directed to KOS cultural interoperability apart from a few articles written by authors from KO community (Dahlberg, &. Siwek, 1995; Miller 2002; Hudon, 2005; Fox & al. 1998; Favier & Mustafa El Hadi 2013). The issue is explored mainly in other discourse communities: in organizations (Bekkers 2005; Kousouris & al. 2011; Cayir & al., 2008) military and diplomatic domains (Clément, 2007, cited by Favier & Mustafa El Hadi (2013); Wimmer & al., 2008; Winslow & al. 2001). However there is a lack of general theoretical framework for considering these developments

KOS in general have long been used as a means of information exchange and has functioned as de-facto standard in knowledge organization and resource discovery: KOS are based on widely accepted and known scientific and educational consensus; as they are conceptual schemes they are language independent, widely used and available in many languages. Bibliographic classifications have some advantages over other controlled vocabularies such as thesauri, descriptor systems and subject heading lists, in that the notational encoding acts as a control device which is linguistically neutral. Retrieval problems such as the proliferation of synonyms, ambiguity of meaning, and other natural language factors can be avoided, making the exchange of information and cross-searching in a multi-lingual or multinational

[5]Metadata can be divided into four main categories not mutually exclusive and there is some degree of overlap between them according to its origins (Broughton 2010:50) professionally generated metadata, created by cataloguers and indexers using conventional controlled vocabularies

- author generated metadata, usually in the form of keywords added to the preliminaries of a paper
- automatically generated metadata, derived from the resource itself using text mining or natural language processing tools, or harvested from the resource's existing metadata
- user-generated metadata, attached by end users, usually on an informal and individual basis, but increasingly as a result of co-operative projects.

environment much easier. Classification process is at the core of all KOS whatever their structure and content is. For Brian Vickery " Classification manifests itself, not only explicitly in the form of classification schemes, but also implicitly in other kinds of information languages which signal conceptual relationships between subjects, most notably thesauri and subject heading lists", (Coates, 1988: 222). We believe that combining the two types of KOS (classifications & language-based KOS such multilingual thesauri and vocabularies) could guarantee full interoperability.

Moreover, KOS are semantically rich terminology tools, traditionally used in library and bibliographic domains to enable information discovery - they are ubiquitous pathways to information in hundreds of thousands of document collections digital libraries, digital collections, and digital repositories. The full integration of KOS in digital environments would facilitate integration of the large corpora of recorded knowledge in heritage institutions and digitally-born resources on the Web. But most importantly, KOS can improve access to document contents in digital or hybrid collections and are indispensible in heterogeneous collections of documents (text, images, audio recordings, datasets, (Slavic, 2011; 2013)).

An overview of KOS has been presented by Hodge (2000), who grouped them into three general categories: (1) term lists, which emphasize lists of terms, often with definitions; (2) classifications and categories, which emphasize the creation of subject sets; and (3) relationship lists, which emphasize the connections between terms and concepts.

Many authors pointed out KOS heterogeneity and the challenges that their compatibility[6] and interpretability represent. Heterogeneity characterizing Knowledge Organization Systems is embedded in their different structures and purposes even if they may vary in their structure, semantic, lexical, and notation or entry features (Iyer & Giguere, 1995, cited in Zeng & al. 2004). For example, they may cover different subject domains or have different scope and coverage; they may have semantic differences that are caused by variations in conceptual structuring (see also Soergel 2010)[7]; their levels of specificity and the use of terminology may vary; and the syntactic

[6] Early endeavors (1995) had been undertaken by ISKO to pave ways for making the KOS compatible, see Dahlberg, Siwek, (1995)

[7] Soergel (2010: 12-36) suggested an approach to mapping between KOS including ontologies, classifications taxonomies thesauri and natural language based on deep semantics. One of the most interesting features of his work, because it is particularly relevant to our investigation, are the examples driven from Cross-language mapping and shades of meaning

features, such as the word order of terms and the use of inverted headings, may be different. Such incompatibilities have presented problems in most mapping efforts. Initiatives for developing concordances or translation between a thesaurus and a classification or among various systems constructed on different principles are extremely challenging tasks. This is especially true when the target system has a higher level of specificity than the source system(s) (for a complete report see Zeng & Chan 2004: 386). But the most important distinction between the different kinds of KOS seems to be the diversity of semantic relations being displayed as noted by Green (2002: 15). In traditional classification systems, hierarchical relations and the relationship between synonyms and homonyms are the most important features[8].

4. Defining Interoperability and its Layers

The term *interoperability* has many meanings, including the notions of communication, exchange, cooperation, and sharing of resources between systems. In fact, the essence of interoperability is that it is a *relationship* between *systems*, where each relationship is a manner of communication, exchange, cooperation and sharing (Carney & al. 2005).

This concept has an overall definition and more specific ones related to the different contexts in which it can be considered. Through the years, interoperability has tended to develop a broader scope and to relate to many contexts. I will now review the different definitions of the term and explore its scope. Paul Miller in his article “What is Interoperability? What Is It and Why Should I Want It?” (2002), noticed that the difficulty to define this concept: “Although not a perfect [definition] this comes closest to capturing my view of interoperability, and is informed by conversations over the past few years with people too numerous to name, all of whom have helped to refine the

[8] Green (2002, 15) pointed out the reasons of their heterogeneity and listed some of the solutions adopted to overcome this difficulty: “Their number exacerbates a problem that arises naturally from their individuality, that is, (some degree of) incompatibility. Responding to this situation, we have also developed a number of tools that show correspondences across classificatory systems, some of the same type, some of different types. Examples include: conversion tables between the Library of Congress Classification and Dewey Decimal Classification schemes correspondences within Library o f Congress Subject Headings to equivalent classes within the Library of Congress Classification ; cross-language equivalences built into multilingual thesauri; the mapping of dozens of medical vocabularies to the Metathesaurus of the Unified Medical Language System (and mappings between versions of the WordNet lexical database But even where mappings across classificatory systems exist, we have no guarantee that the end result will lead to a unified semantic content and structure”

ideas behind these words: *to be interoperable,* one should actively be engaged in the ongoing process of ensuring that the systems, procedures and culture of an organization are managed in such a way as to maximize opportunities for exchange and re-use of information, whether internally or externally".

This general definition can be also compared to its scope in the enterprise context while referring at the same time to technical interoperability in its large sense: "Interoperability is the ability of disparate and diverse organizations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organizations via the business processes they support, by means of the exchange of data between their respective information and communication technology (ICT) systems", in order to cooperate seamlessly, in an automated manner, in depth of time for a common objective" (Koussouris & al. 2011:1).

More definitions in the enterprise context where it is considered to be a major issue are suggested by the authors: "Enterprise Interoperability is perceived as a capacity of two or more enterprises, including all the systems within their boundaries and the external systems that they utilize or are affected by, in order to cooperate seamlessly, in an automated manner in depth of time for a common objective" (Koussouris & al. 2011: 1). The role of ICT in achieving interoperability in enterprise is also highlighted in Charalabidis & al. (2008, cited by Koussouris & al. 2011) the authors consider enterprise Interoperability as a "a field of activity with the aim to improve the manner in which enterprises, by means of Information and Communications Technologies (ICT), interoperate with other enterprises, organizations, or with other business units of the same enterprise, in order to conduct their business ». More interesting is their view of interoperability and its obvious links with semantic interoperability in KOS: "Consideration of the terms "Enterprise" and "Business Transaction" and decomposition of the enterprise concept in its major ingredients (e.g. people, assets, processes, knowledge, etc.); (..). Identification of the key EI challenges, as documented in the EI Research Roadmaps, (…); definition of a common EI taxonomy glossary in order to ensure common understanding of the key underlying terms. During this step, four different granularity levels for the EI taxonomy have been defined based on the prerequisites they require"

In the military and diplomatic fields, interoperability is defined in the US, NATO and Australia as "the ability of systems, units, or forces to provide services to and

accept services from other systems, units, or forces and to use the services so exchanged to enable them to operate effectively together". (Koussouris & al. 2001).

Donna Winslow put forward the reasons for considering cultural interoperability in military and diplomatic contexts. In a report submitted to the UN in 2000[9] she gives the following definition: " In order to promote civil military cooperation, it is important to understand some of the difficulties that can arise in peace operations between the members of these communities. (…) I will explore some of the tensions that can arise between the military and relief agencies. I have identified five possible points of tension to be found in peace operations, which I have been calling, a "cultural interoperability" model. These points of tension are related to organizational differences in terms of:

1. organizational structure and culture,
2. tasks and ways of accomplishing them,
3. definitions of success and time frames,
4. abilities to exert influence and control information,
5. control of resources. (see also with reference to a previous study", (Winslow & al. 2001).

4.1. Interoperability in Knowledge Organization Systems

Closer to our concerns are the definitions and scope of the term suggested by Fox (Fox & al. 1998). The authors focused on interoperability in the context of Digital libraries thus identifying different layers likewise Miller (2002) who pointed out the different "flavors" of interoperability: " Linking multiple systems in a useful way requires various levels of interoperability: Technical interoperability, which is concerned with

[9] The report proposes that the UN should be able to deploy 5000 troops "as a brigade formation, not as a collection of battalions that are unfamiliar with one another's doctrine, leadership and operational practices." (UN 2000: 19) This is an important point. Outside of NATO militaries there is little commonality therefore UN operations run into problems due to a lack of compatibility or as I call it a lack of "cultural interoperability." This can affect the UN's ability to assure human security. In order to examine possible points of tension, which can arise in peace operations due to a lack of cultural compatibility between the various militaries in the UN operation in Bosnia, shad to develop a measure of "cultural Interoperability" . In an article wrote with her husband Colonel Peer Everts (Winslow and Everts 2001) she developed the concept of a cultural interoperability model. In UNPROFOR there was a great diversity in the organization of the operation and as my colleagues in COM would remind us, organizational studies show that diversity in organizations leads to increased communication and co-ordination problems and thus to potentially decreased organizational performance. We also demonstrated that one reason for the success of NATO in IFOR and SFOR was a common military culture shared between NATO participants in the operation. This commonality had been developed over fifty years of training and working together. Thus experience in careful planning, clear hierarchies and strong discipline ensured co-operation and co-ordination in diversified structures typical of multinational operations (Winslow & al. 2001).

hardware, networks, data types, and application compatibilities and protocols; Informational interoperability, which addresses content scope, language, metadata, naming conventions, semantics, and user Interfaces; Social interoperability which is concerned with personal and organizational rights and responsibilities" (…) It is clear that all levels of interoperability, including the social-political, must be addressed for a more tightly coupled global digital library to emerge".

Paul Miller (2002) gave a detailed account of the concept and identified different levels of interoperability calling them "Flavors of interoperability". He noticed the following: "Given such a wide scope within the suggested definition, it becomes useful to further subdivide the notion of interoperability". In his broad definition, Miller distinguished more layers in addition to the most common ones (semantic, syntactic and technical interoperability). He added political and human interoperability which has to do with the decision to make resources more widely available and has implications for the organizations: the inter-community interoperability; the legal interoperability and the international interoperability to which are related cultural issues. This involves usage practices, expectations, and requirements that vary from country to country. The layers added by Miller are quite relevant for anyone intending to integrate data from different locations.

We believe that semantic interoperability is the most important layer because it has an important impact on cultural interoperability: different types or levels of search and data interoperability, such as syntactic, functional and semantic, have been identified (Moen, 2001:2). He argues, "While syntactic and functional levels may focus on protocols used for information retrieval and communication systems, the semantic level concerns the understandings and meanings of interchanged data (…). In the hierarchy of challenges for effective interoperability, semantic interoperability is regarded as a grand-challenge research area for achieving high-quality interoperability". We will now see how semantic interoperability impacts cultural interoperability.

4.2. Semantic Interoperability

Semantic interoperability refers to the meaning of information to its human users, as opposed to the simple physical transfer of data. Interoperability at this level can fail if different users, or groups of users, use different terms for the similar

concepts, or use similar terms to mean different things. Friesen (2002: 105) argues that "semantic interoperability is tied directly to communities of practice, and to the negotiation of meaning that occur within them." Because meanings assigned by people and groups can vary over time and in subtle ways, semantic interoperability cannot be as easily ensured as technical interoperability but it can be improved. Attention must be paid to who the potential users of information systems are and what are the language that are necessary to communicate with them; there should be an agreement on the standard thesauruses and lists of terms that are to be used in metadata systems and there should be a consistent use of nationally existing coding systems [10]

5. Review of Trends in Establishing Semantic Interoperability between Knowledge Organization Systems

The main foci of recent research activity into establishing semantic interoperability will be reviewed in this section with a special focus on multilingual thesauri.

The use of KOS to support more effective organization and retrieval of digital resources is not in itself a new concept. Web development in the mid-nineties was marked by numerous research projects building quality subject gateways and portals using classification and thesauri for browsing. The idea of using classification as a 'switching language' or 'pivot' to map indexing languages for the purpose of information integration and exchange was widely discussed as early as the 1970s (Marcella & Newton, 1995, cited in DIGIKO, 2013). This trend was especially strong in the UK where a number of SGs were developed within the Electronic Libraries Programme (eLib) (Kerr, 2009). As manual indexing has become increasingly difficult towards the end of the 1990s a number of projects explored automatic classification of web resources (GERHARD, SCORPION, CORC, Renardus). The evolution of the Web, from purely academic to commercial, business and social coincided with a significant improvement in technology. Web portal developers started to be primarily interested in the benefits of concept organization into practical and purposeful categories that their

[10] Guide for *Australian Flexible Learning Framework for the National Vocational Education and Training (VET) System*, produced by the Sharing Knowledge project, an initiative within the *2000-2004.*

users/customers could easily combine or independently navigate. Subject orientated, simple hierarchical structures on gateways and portals have started to be combined with, or replaced by, a faceted organization of object/subject properties and attributes. These applications led to the development of data formats and tools for the management of faceted vocabularies and their use in a Web interface (La Barre, 2006; Giess, Wild & McMahon, 2008, cited in DIGIKO, 2013). The wider exposure of tools such as thesauri and classification systems created a need for better understanding of the theory and principles underlying these tools. For instance, the lack of understanding of more sophisticated knowledge organization systems such as faceted classification resulted in web and other vocabulary standards being weak in supporting these systems (Slavic 2008; 2011, cited in DIGIKO, 2013).

Classifications are particularly well suited for use as a central mapping spine in connecting the vocabularies of different languages and seem to give visible results quickly and easily. This idea was explored by OCLC in terminology services (Vizine-Goetz et al. 2004). With the emergence of MARC formats for classification data, ISO standards for structured vocabularies and especially since the creation of SKOS which initiated most of copyright protected vocabularies to be released as linked data on the Open Web we can see a start of a new stage of KOS use in the digital environment. Over past five years many of these vocabularies were published as linked data on the Web and the main obstacle for their wider and quicker deployment is lack of expertise. It makes economic sense that the knowledge and experience associated with these tools are made more universally accessible and usable by both peoples and programs across sectors and domains. This is the first step towards the increasing participation and interaction of different knowledge communities in improving KOS terminology and enhance integration and linking of knowledge landscape. Moreover, publishing of KOS on the Web opens many possibilities of exploitation of KOS in many fields of activities in which they were not so easily available. Knowledge organization as a field and KOS as tools can thus be able to operate within the whole universe of recorded knowledge thus providing a unique semantic layer for knowledge discovery. Dextre Clarke & Zeng (2012: 23) showed how is SKOS vital in such endeavors: “The W3C Recommendations SKOS is designed to support publications of vocabularies such as thesauri on the Web. And at its heart is a data model that explicitly distinguishes between concepts and labels to represent concepts. (…). In addition to all that is

conveyed by SKOS Core for relationships between concepts, the extension provides additional support for identifying, describing, and linking lexical entities", Dextre Clarke & Zeng (2012: 23), (see also Dextre Clarke, 2011)

To this end, over the past ten years a number of W3C workgroups in the area of Semantic Web development have showing a keen interest in knowledge organization systems traditionally used in the bibliographic domain (thesauri and classifications) and their power in capturing and communicating meaning. This is especially the case with W3C development groups in the area of vocabularies, web ontology standards, linked data and more recently the W3C Library Linked Data Incubator Group. This sparked research activities, projects and forums such as Networked Knowledge Organization Systems (NKOS) that promote collaboration between bibliographic and other domains in the area of subject retrieval.

The wider exposure of tools such as thesauri and classification systems created a need for better understanding of the theory and principles underlying these tools. For instance, the lack of understanding of more sophisticated knowledge organization systems such as faceted classification resulted in web and other vocabulary standards being weak in supporting these systems (Slavic 2008). Vanda Broughton (2010: 50) highlighted the proximity of SKOS with the faceted classifications and their relevance for the semantic Web: "In order to interact fully with the Semantic Web, a faceted classification must be made visible (Slavic, 2008, cited by Broughton, 2010), and all of the elements that make faceted KOS particularly useful for online organization and retrieval must be visible too. Some existing Semantic Web applications, such as Simple Knowledge Organization System (SKOS) exhibit features similar to the faceted KOS, and indicate that faceted KOS may provide a model for the elements in Semantic Web formats, as well as the terminologies to inhabit them".

In Mustafa El Hadi (2013), I investigated the ways in which universal classifications, as language independent concept schemes, can assist humans and computers in structuring and presenting information and formulating queries. A wider application of universal knowledge classification in digital humanities may contribute not only to better resource discovery but could also help in improving terminological and research tools for knowledge transfer between humanities and other fields of knowledge.

5.1. Mapping between Multilingual Vocabularies

Establishing of equivalence is essential when creating multilingual subject vocabularies or merging multiple vocabularies. According to the assumption that all languages are equal in a concordance, there exists a question of whether views of a particular culture that are expressed through a controlled vocabulary or a classification can be appropriately transferred during the mapping (Zeng & al. 2004)

The ideal match is the establishing the “one-to-one” relationships between terms in different vocabularies and different languages, but this often proves elusive (Zeng & Chan 2004). To map vocabularies or create multilingual or multi-disciplinary vocabularies other difficulties crop up when trying to establish equivalence between these KOS. Different linguistic expressions for the same concept, different degrees of specificity and polysemic terms are hindering these attempts.

The ISO 5964 standard specifies the following types of equivalence relationships: exact equivalence; inexact equivalence; partial equivalence; single-to-multiple term equivalence; non-equivalence (ISO, 1985). Similar attempts at defining term equivalence reported for by (Zeng & Chan 2004) include the degrees of equivalence suggested by Riesthuis (2003) who has defined them as follows: exact equivalence, single-to-multiple equivalence, non-equivalence, loan terms, and inexact equivalence.

Although there are complex challenges to achieve semantic interoperability, many efforts have been made into this direction in order to support high quality information exchange and integration across communities (Chen, 1999; Chung & Moens 2001; Zeng & Clarke, Hoolande & al. 2013; Doerr, 2001. Chung & Moens (2004) identified two primary areas in semantic interoperability research: data attributes and data values. The data attribute area addresses questions as to the names, labels, semantics, and granularity of metadata scheme elements and database fields. The data value area addresses the data or information provided in an element or database field. Their research is based on Doerr (2001) approached the semantic interoperability issue with a concept-based mapping between numerous vocabularies for subject retrieval illustrating the relationship between two dimensions—data attribute and data value in the context of semantic interoperability. He noted that as the semantic interoperability of the two dimensions increases, the overall semantic interoperability increases proportionally. To improve the capability for information

exchange or integration, and to improve the search capacity across these resources, it is essential to enhance interoperability of systems and data. This involves addressing not only interoperability at the syntactic and functional levels but at the semantic level (Moen, 2001). Various information services such as federated searching, harvesting metadata, and gathering are often dependent on the capability and quality of semantic interoperability across different collections, systems, domains, and communities (Arms et al., 2002; Chun & Moens 2004).

Chung & Moens (2004) highlighted the need for semantic interoperability to meet users' expectations of discovering relevant networked information resources not only via searches across heterogeneous as well as within homogeneous systems and collections (Chen, 1999; Tennant, 2001, cited by Chung & Moens (2004). This is the reason why semantic interoperability is considered as in the "hierarchy of challenges for effective interoperability", Chun & Moens (2004:2). We believe therefore that the most important step towards effective interoperability is conditioned by the semantic layer

To make different KOS compatible, many efforts have focused on term or concept mapping. Both intellectual and machine-based efforts to map among vocabularies in the online era have been reported for by Zeng and Chan who categorized them in the following groups (Zeng, 1992; Zeng & Chan 2004). The authors reported on many initiatives aiming at unifying KOS in an integrated source or interface. Efforts at establishing interoperability among KOS in different languages and in different structures can lead to the integration, mapping, and creation of knowledge organization systems for information sharing in a networked environment as Zeng & Chan (2004) indicated. These projects vary in both the targets for mapping and in the methods used (for an extensive description of methods, KOS categories, (see Zeng & Chan 2004). The authors described the projects in terms of the languages and structures involved. The efforts in establishing interoperability among existing KOS in different languages are examined and grouped in two categories: the first, consisting of MACS and Merimee, attempts to establish links among the vocabularies or authority files involved. The second, consisting of the remaining examples, utilizes a switching mechanism among the vocabularies involved. Ten projects had have been reported for (Zeng & Chan (2004: 379-380). As for establishing interoperability among different existing KOS, the projects described have been categorized by the authors according

to the structural types that the involved KOS possess. For an extensive description of the projects see Zeng & Chan (2004: 380-381). Regarding the methodological analysis, they analyzed both the conventional and the new ones used in mapping such as derivation/modeling; translation/adaptation; direct mapping; co-occurrence mapping; switching; linking through a temporary union list; linking through a thesaurus server protocol. They also analyzed methods used in link storage and management. They comprise: authority records, concordances, semantic network and lexical database (2004:386).

In their conclusion the authors confirmed the following: interoperability of KOS is an unavoidable issue in today's networked environment. In addition to efforts to achieve interoperability among existing KOS for specific collection(s) and project(s), collection-independent KOS, especially vocabulary knowledge bases, are highly desirable. The need for reconciling different subject vocabularies in networked environments is indisputable. Results from recent efforts to achieve interoperability among different vocabularies are encouraging. These projects have identified and experimented with a variety of methods. The question remains as stated by them (2004-386): "Have we fully exploited technological capabilities in our efforts to improve subject access to the myriad resources now available in the networked environment?" Their analysis and findings are extremely valuable for our investigation.

In a recent paper, Hoolande (Hoolande & al. 2013: 464) explored the feasibility of using subject vocabularies as a linking hub to the Semantic Web. They noted that the semantic enrichment and integration of heterogeneous collections can be facilitated by using subject vocabulary for cross linking between collections, since major classifications and thesauri (*e.g.*, Library of Congress Subject Headings [LCSH], Arts and Architecture Thesaurus [AAT], DDC, Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié [RAMEAU]) have been made available following Linked Data principles. Reusing these established terms for indexing cultural heritage resources represents in their view a big potential for LAM (Library, Museums and Archives). For this purpose they examined the feasibility of using subject vocabularies as a linking hub to the Semantic Web in advance of such efforts. They

adopted a semi-automated vocabulary alignment technique used earlier by Doerr (2001), Isaac et al. (2008)[11].

So here is, a non exhaustive summary, compiled from the various methodologies used in the literature, that are used to achieve semantic interoperability

1) mapping between numerous vocabularies (Doerr, 2001)

2) concept-based mapping between numerous vocabularies for subject retrieval Doerr (2001)

3) classifications used as a central mapping spine in connecting the vocabularies of different languages being used (Marcella & Newton, 1995; Vizine & al. 2004)

4) using subject vocabulary for cross linking between collections using subject vocabularies as a linking hub to the Semantic Web (Hoolande & al. 2013)

5) establishing a variety of links among the vocabularies or authority files involved (Zeng & Chan, 2004)

6) using a switching mechanism among the vocabularies (Zeng & Chan, 2004)

5.2. Multilingual Mapping: the Case of Multilingual Thesaurus

Why do I focus on thesauri? The reason is that this tool represents a category of KOS built on natural language. They are increasingly multi-lingual and capable of accommodating complex cultural demands. The major challenge of multi-lingual thesauri, making controlled vocabulary work across language boundaries (Smiraglia 2012). Examining this KOS category is one of the key elements for considering cultural interoperability. This idea will be developed in section 7.

Heterogeneity between different thesauri is attributed to the following, as identified in Doerr (1996; 2001) and Dachelet 1997: different word uses, due to different natural languages; different coverage, due to different states of development, different scope and varying user needs; different semantics, due to different conceptualizations.

[11] Doerr (2001) and Isaac & al. (2008) İdentified four general approaches toward vocabulary reconciliation or alignment: (1) lexical alignment techniques, (2) structural alignment, (3) extensional alignment, and (4) alignment using background knowledge (Hoolande & al. 2013: 465-466)

This will occurs typically between thesauri from different languages, but it may also be due to different aspects of classification; different semantic relations are often due to the enforcement of mono-hierarchies, but they can also be due to different classification aspects (Doerr, 2001).

Doer (2001) observed that any kind of relations between terms from thesauri in different natural languages are referred to as translations. In his opinion, translation *per se* differs from concept-based mapping and cross-concordances in significant ways. To this end he preferred referring to the typology suggested by Roland Dachelet (1997, cited by Doerr, 2001). This author proposed distinctions between different kinds of multilingual thesauri, defining “translated thesauri” as thesauri where each concept is optimally interpreted in words of another or multiple languages, to allow speakers of those languages to understand better and use the concepts of this thesaurus more effectively. He noted that such translations are in general not established indexing terms of the target language; he then defined correlated thesauri as an aggregate of multiple thesauri consisting of established indexing terms (concepts) of the respective user communities and a set of concept-based mappings between the concepts from the different thesauri of that aggregate; then the Interlingua Thesaurus was defined as a thesaurus made of concepts created by fusing each cluster of similar concepts from different social groups into a new concept. The Interlingua provides a sharing of concepts between social groups. It may not contain any of the original concepts of any user group but it contains a set of compromises to remove interpretational differences. Its concepts may again be translated and correlated to other thesauri.

Zeng & Chan (2004) reviewed a set of methodological options for multilingual mapping. In order to meet users’ requirements in networked environments, a major decision facing those attempting to establish interoperability among knowledge organization systems is the choice of an appropriate method. The first question is whether to integrate, map, or create a system. The possible options are similar to those suggested by Riesthuis (2003, quoted in Zeng & Chan (2004)) with regard to different approaches in creating multilingual thesauri: translation, merging, and creating from scratch.

Dagobert Soregel[12] summarized the main limitations of existing KOS taking as an example the AGROVOC thesaurus. This can be seen as complementary to Hudon (1997, 2005; and Doerr 2001; Zeng & al. 2004) He mentioned successively:

-Lack of conceptual abstraction: thesauri and other traditional KOSs are collections of terms (generic or domain-specific), ordered in a poly-hierarchic lattice structure or a mono-hierarchic tree structure and interlinked with some very broad and basic relationships.

-Limited semantic coverage: most thesauri do not differentiate concepts into types (such as *living organism*, *substance*, or *process*) and have a very limited set of relationships between concepts, distinguishing only between hierarchical relationships, i.e. NT/BT, and associative relationships, i.e. RT.

-The concept relations provided by most thesauri force all relations into two broad categories, hierarchical and associative.

-Lack of consistency: since the relationships in thesauri lack precise semantics, they are applied inconsistently, both creating ambiguity in the interpretation of the relationships and resulting in an overall internal semantic structure that is irregular and unpredictable..

-Limited automated processing: traditionally thesauri were designed for indexing and query formulation by human beings and not for automated processing. The ambiguous semantics that characterizes many thesauri makes them unsuitable for automated processing.

In a survey on multilingual thesauri, Michèle Hudon (1997: 89-90) noticed the following problems:

The stretching of a language to make it fit a foreign conceptual structure to the point where it becomes barely recognizable to its own speakers; The transferring of a whole conceptual structure from one culture to another, no matter whether it is appropriate; and The literal translation of terms from the source language into meaningless expressions in the target language.

According to Hudon (1997-90): “true equality for all languages in a multilingual thesaurus has a better chance of being achieved if the following global requirements

[12] Dagobert Soergel, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer and Stephen Katz “Reengineering Thesauri for New Applications: the AGROVOC example (available at http://www.fao.org/docrep/008/af234e/af234e09.htm#TopOfPage)

are met: the thesaurus should be built within a semi-centralized administrative structure, with representatives of each language/culture on the decision-making team; all linguistic versions of the thesaurus should be developed simultaneously from the ground up; the thesaurus designers should be native speakers of the language in which they work, with a good knowledge of the other languages involved; distinct term banks should be built independently or each language with terms found in source language documents; identity and symmetry of structures are not required across the various linguistic versions of the thesaurus, and single-to-multiple equivalence, “orphans", and variations in hierarchies, etc. are allowed; the use of neologisms should be very restricted if allowed at all; thesaurus development software which allows for nonidentity of descriptor records and for rotation of source and target languages should be used; physically separate displays for each language represented should be produced”.

7. Cultural Interoperability

The concept of Culture is disciplinary and describes in general the various phenomena that make up the collective beliefs and activities of a certain group of people. Discussions of culture refer generally to shared values, history, language, collective memory, social attitudes, preferences and practices (Beghtol 2002a:509).

In its early definitions cultural interoperability is related to the military and diplomatic fields which are considered as organizational bodies in the way we mentioned above. The term can be traced back[13] in Winslow and Everts writings (2001). "In an article I wrote with my husband Colonel Peer Everts (2001:19) I developed something called a “cultural interoperability model” in order to examine possible points of tension, which can arise in peace operations due to a lack of cultural compatibility between the various militaries in the UN operation in Bosnia. In UNPROFOR there was a great diversity in the organization of the operation and as my colleagues in COM would remind us, organizational studies show that diversity in organizations leads to increased communication and co-ordination problems and thus to potentially decreased organizational performance.”

Sotirios Koussouris, Fenareti Lampathaki, Spiros Mouzakitis, Yannis Charalabidis, and John Psarraas suggested a broader definition of cultural interoperability. Their definition applies to the enterprise and business contexts and

[13] According to King (2006), Bernard Boene was the first author who coined this term

should perhaps be seen as strongly limited to these contexts: “Interoperability between organizations, people and enterprises systems that have different languages and different cultural aspects such as Politics, Religion, regional Art, Traditions and Social Customs defines the concept of *Cultural Interoperability*. As such, Cultural Interoperability is the degree to which knowledge and information is anchored to a unified model of meaning across cultures. Enterprise systems that take into consideration Cultural Interoperability aspects can be used by transnational groups in different languages and cultures with the same domain of interest in a cost-effective and efficient manner. Cultural interoperability mechanisms are based on the assumption that both high-level and low-level layers of Enterprise reflect culture and that the linguistic encoding of knowledge and information is therefore culturally biased. These interoperability mechanisms and considerations address the ability of enterprises to understand and co-manage context from any source and of any kind, therefore realizing the cooperation between enterprises with major cultural differences”

7.1. Cultural interoperability in Knowledge Organization (KO)

As I mentioned in the introduction a few articles have dealt with cultural interoperability in the Knowledge Organization community. In a recent article (Favier & Mustafa El Hadi, 2013) we examined two possible dimensions of semantic interoperability: one based on KOS Translatability, the other one, on navigating within their diversity, exploiting links between heterogeneous KOS for information discovery and retrieval purposes[14]. In other words we tested in the Visual Ctalaog (Papy, 2013) what is called the approach of a “one-stop” seamless search process (see also Zeng & Chan 2004), when entering one query the user can retrieve results from various heterogeneous KOS. This assumption lead us to put forward the following hypothesis: semantic interoperability is the most fundamental step towards cultural interoperability.

7.2 The concept of Translatability

[14] This is a major issue when considering convergence of information sources, the Web, digital libraries, archives, museums and repositories and the need for multidisciplinary and multicultural information. Web scale discovery tools improving search and retrieval such as OCLC or EBSCO federated search shows the directions to follow today. This technology allows the simultaneous search of multiple searchable resources. A user makes a single query request which is distributed to the search engines participating in the federation. The federated search then aggregates the results that are received from the search engines for presentation to the user.

Our hypothesis is built on translation possibilities between different languages, called translatability in linguistic communities. I will examine here the scope of the concept of “translatability ” of and how it applies to KOS. From the translation community, Raquel de Pedro (1999: p. 546-559) accounted for two points of view when considering translatability: the universalist one and the monadist one. Supporters of the former approach claim that the existence of linguistic universals ensure translatability. Those who endorse the latter approach maintain that each linguistic community interprets reality in its own particular way and the universalist approach jeopardizes translatability . The approach, that has become known as the Sapir-Whorf hypothesis is generally not applied in its strongest form, since this would imply the impossibility of effective communication between the members of different linguistic communities (see also Green, 2002).

From Knowledge Organization community, Rebecca Green showed how translatability is linked to semantic and lexical universals. “The extent to which we can satisfactorily translate between human languages is, it would seem, a good measure of the degree to which natural language incorporates semantic and lexical universals. (…) In enumerating conceptual universals that must be accommodated by any comprehensive knowledge organization or knowledge representation system, we may productively look to the semantic and lexical universals found in natural languages” (Green 2002: 20).

Green, Bean, & Hudon (2002: 311) examined whether a concept’s hierarchical level affects the likelihood of its universality across schemes for knowledge representation and knowledge organization. The authors drew empirical data on equivalents from a bilingual thesaurus, a pair of biomedical vocabularies, and two ontologies. Their findings show that “conceptual equivalence across resources occurs significantly more often at the basic level than at subordinate or superordinate levels. Attempts to integrate knowledge representation or knowledge organization tools should concentrate on establishing equivalences at the basic level. The implications of semantic and lexical universals survey showed the existence of conceptual universals in knowledge organization and knowledge representation systems” (for a complete account of this survey see (Green 2002; Green et al. 2002). It appears to the authors that there are dozens of conceptual units (e.g., sun, bad, think) that are more or less universally apprehended. There are also some numbers of conceptual abstractions

(e.g., predicate-argument structure, countability, animacy, hierarchy) that are used universally in structuring our conceptual systems (…) where specific conceptual units are shared, we should construct cross-system mappings.

What would justify characterizing some conceptual content as a necessary feature of all human language? Green referred to Immler (1991) who took up a variety of meanings purported to exist in all human (he chose 18 languages) — for example, the distinctions between light and dark and between alive and dead— and then suggests “no human society could afford not to be able to express these ideas in its language” (pp. 39-40), Cited by Green (2002: 18). “cognitive contents can be universal across all cultures and all languages— they will be universal to the extent that the real world around us , which is the same for humans all over the world , is the object of cognition” argues Immler.

This idea resonates with the assumption according to which, in the communication process, as well as in the case of interoperability, perception and cognition play a leading role in building, structuring, and disseminating human information. Human communication begins when someone wants to transmit information in their mind to someone else. As such, cognitive models are basic elements of human-to-human communication (Brodeur et al. 2003)

Green, Bean, & Hudon, (2002: 311-312), in their seminal paper suggested that “conceptual units shared by many languages are more likely to occur at the basic level than at hierarchical levels above or below the basic level” suggest that the “best we can do in integrating two knowledge-based schemes is to construct such crosswalks as are truly appropriate and then rely on the individual structures of the schemes being “integrated.” Knowledge of these conceptual universals can at least facilitate the very considerable work that knowledge integration will always require”.

The assumption that meanings are given in a language, and that subjects can be constructed by a universal set of semantic universals given by natural languages has been questioned by Hjørland (2003: 109) when he posited that: “The basic assumptions and attitudes have been that the basic units of KO are semantic relations between concepts. Such semantic relations cannot primarily be established by universalistic assumptions, but much primarily be understood as domain specific, as uncovered by (and constructed by) scientific disciplines”. But we believe however, that our assumption makes sense while attempting to build general framework for

describing what cultural interoperability in KOS is. Linking the two points of view (translation theory [15] and techniques) and (KO theoretical findings) represents another avenue available for research in semantic and cultural interoperability.

7.3. Fundamentals of cultural interoperability for KOS

An analysis of the literature demonstrates that in order to build a theoretical framework to account for cultural interoperability in KOS some basic concepts are to be conveyed/ brought up. I will suggest first considering the concept of "universals in KO" as a basis for considering both semantic and cultural interoperability

7.4. From semantic Interoperability to Cultural Interoperability

Beghtol (2002a : 507) wrote: "new technologies have made the increased globalization of information resources and services possible. In this situation, it is ethically and intellectually beneficial to protect cultural and information diversity". She analyzed in her study the problems of creating ethically and globally accessible and culturally acceptable knowledge representation and organization systems and foundation principles for the ethical treatment of different cultures established on the basis of the United Nations Universal Declaration of Human Rights. From an ethical point of view she mentioned the need for a global and local access to information in any language, available at any location at any time and for any purpose for any individual, culture, ethnic groups or domain. She suggested that the concept of "cultural hospitality" can act as a theoretical framework for the ethical warrant of KOS.

The two concepts, "Cultural Warrant", and "Cultural Hospitality" are relevant for exploring cultural interoperability had been defined in (Beghtol (2002b: 45). Cultural warrant postulates that: "a knowledge organization system is more likely to be useful and appropriate for those who are members of a culture and that it is less likely to be useful and appropriate for those who belong to some different culture, at whatever level of society that culture or domain may reside. Thus, a knowledge organization system

[15] In his own experience as a translator made von Humboldt perceive the need for translation: "translation is one of the most necessary tasks of any literature" (in Schulte & Biguenet, 1992: 56, quoted by Pedro). According to his hypothesis, the structural differences which exist between languages are no obstacle for translation. The reason that von Humboldt proposes to explain this is that each linguistic community has a potential of expression which can generate resources for verbalizing every extra-linguistic area, even those which go beyond its own social and cultural experience

that is appropriate for the elements of one culture may not recognize elements that are highly important for some other culture, and such exclusions pose problems because we need to integrate knowledge across cultural, geographic, and linguistic boundaries. Important tensions thus exist between culturally specific knowledge organization systems and the need to provide systems that are both globally accessible and culturally appropriate. In general, knowledge organization systems for global usage would need to incorporate all the various syntactic and semantic foundations of any and all of the world's cultures, and the creators of knowledge organization systems need to create techniques for poly-cultural information retrieval. Since the foundations of different cultures may be in conflict with each other, we need to develop theories and techniques for incorporating any possible cultural assumptions that might be used for information retrieval"

The concept of "Cultural hospitality" complements and extends the principles of cultural warrant. Both concepts are essential for a theoretical framework for KOS that will privilege the needs and existence of different cultures, whether they are national, ethnic, domain or disciplinary cultures: "The concept of "Cultural hospitality" has long been established as one of the desiderata for the notation of bibliographic classification systems. Hospitality is the ability of a notation to admit new concepts appropriately and to accommodate them in the correct relationships with other concepts. To develop a new theoretical framework for knowledge organization systems, it is helpful to expand the concept of hospitality by 1) including provisions for hospitality beyond notational issues and 2) broadening hospitality to include different cultures as well as new concepts. These extensions of hospitality provide us with a new concept, "cultural hospitality". This concept implies that knowledge organization systems should be "permeable" to different points of view and different cultural attitudes and practices", (Olson 1996, p. 9 quoted by Beghtol 2002b: 47). Although these two concepts are relevant, however their implementation may not be easy as Beghtol observed. She suggests the need for discussing, assessing and testing its potential for effective implementation.

The considerations sketched above can open a path of research that until now has been largely focused on technical and semantic interoperability. What hinders semantic interoperability and consequently cultural interoperability is that the degree of success achievable in the integration of multiple knowledge representation systems

or knowledge organization schemes constrained by limitations on the universality of human conceptual systems. These limitations have been observed in the case of multilingual thesauri (see above section 5.2.). Natural human languages do not all lexicalize the same set of concepts; nor do they structure (quasi-)equivalent concepts in the same relational patterns (….) As a consequence, even multilingual thesauri designed from the outset from the perspective of multiple languages may routinely include situations where corresponding terms are not truly equivalent (Hudon, 1997, 2005), Green et al. 2002)

It is obvious that the lack of equivalence in mappings across knowledge organizations schemes severely impacts retrieval scenarios. The limitations are due to the nature of concepts at the basic level, where concrete entities are more likely to accept mapping through different languages as suggested in Green (2002: 23) “The basic level concept arose in the context of classifying physical objects. It is not always clear how to apply some of the converging processes noted above, especially those related to perception and function, to abstract concepts, processes, events, and so forth (…) We will assume, however, that the basic level concept does apply to these less concrete contexts and that the several linguistic criteria that tend to converge for concrete entities will likewise converge for non-concrete entities. The confirmation of our hypothesis that full equivalence between tools occurs more often at the basic level than at either superordinate or subordinate levels for both knowledge representation schemes and knowledge organization schemes should illuminate future efforts to build universal tools of subject or semantic access. Specifically, it suggests that crosswalks between such tools should focus on mappings at the basic level, without attempting to impose a comprehensive mapping at all hierarchical levels”

7.5. Domain Analysis as an Element of a General Theoretical Framework for Examining Cultural Interoperability

Domain Analysis in its broad scope can possibly be adapted to a general framework addressing cultural interpretability in KOS. In his paper on “Epistemology of Domain Analysis, in *Cultural Frames of Knowledge”*, Smiragilia (2012: 111-112) noted that “Domain Analysis is at the heart of KO, for without it we would have no ontological matter to constitute our KOS. Culture and activities” (…) The domain is best understood as a unit of analysis for the construction that is a domain is a group with

an ontological base that reveals an underlying teleology, a set of common hypotheses, epistemological consensus on methodological approaches, and social semantics". The concept of "domain" and the scope of Domain analysis I am considering here is based on Hjørland and Albrechtsen (1995, p. 400) hypotheses, extended by Smiragilia (2012); and Tennis (2003).

Hjørland and Albrechtsen (1995, p. 400) define domains as "thought or discourse communities, which are parts of society's division of labor" (..). Smiragilia (2012:111) noted that "A domain can be seen, according to these authors, as a type of discourse community". Domain Analysis definition and scope had been extended and developed by Smiragalia (2012) and Tennis (2003: 191). Tennis considers the definition of "domain" as prior to defining DA definition. In reference to Hjørland and Albrechtsen text, he wrote the following: " (..), the term domain is not the same as discourse community. [They] go on to review the literature that uses the concept of domain under many terms. They cite "speciality/ discipline/domain/environment" (Hjørland and Albrechtsen, 1995 p. 401) as the unit of study. As a consequence, the definition and its boundaries are muddied. It can be noted that for Hjørland (1995, 1998, 2002), it seems more important to define the *domain analytic paradigm* than the object of inquiry, the domain. And it is this problem of definitional muddiness I address here. There are a number of concepts similar to domains in Information Science; like Communities of Practice, and Epistemic Communities. It seems apparent to the casual reader what a domain might be. It could be an area of expertise, a body of literature, or even a system of people and practices working with a common language". Tennis added what he called "the two axes of DA" essential in his view to consider when analyzing a domain: "There are two axes of consideration when analyzing a domain: Areas of Modulation and Degrees of Specialization. Areas of Modulation set the extension of the domain and Degrees of Specialization set the intension. Each of these axes has two parameters. Areas of Modulation must state 1) the totality of what is covered in the domain analysis – the extension and 2) what it is called – its name. The Degrees of Specialization must 1) qualify the domain – state its focus and 2) state where the domain is positioned against other domains – its intersection (2003: 194)

If we sum up the basic elements to consider when building a KOS we will realize that the ingredients for a successful semantic and cultural interoperability between two or more categories of KOS (a classification scheme, vocabulary, thesaurus, ontology,

dealing with the same knowledge-domain are embedded in the very definition of "Domain" (see Tennis, 2003 (191-192) for a an extended definition of Domain, Hjørland 1993, 1995; 2002). A domain has: a boundary, a specific terminology, basic units concepts, terms, semantic relations, classification schemes and a shared ontology.

Hjørland (2002: 422) developed the importance of the role of domain studies in Information Science. While this discipline originates from special librarianship and documentation it lacked from general librarianship. He argues: "the core idea of the specialist library approach may be that information resources should be identified, organized and communicated to serve specific goal". (…). In my opinion even general librarianship has to cope with different domains and may well benefit from considering the domain analytical view. One cannot treat all domains as if they were fundamentally similar, and a theoretical approach to Library and Information Science should consider different discourse communities".

Thesaurus is a domain specific tool built for describing, indexing and retrieval. Theses specific KOS are built on all natural languages, in different cultures and different organizations, sometimes within the same culture or language. Their importance in multilingual contexts is not any more to be proved. Doer (2001:1) wrote: "Thesauri are created in different languages, with different scope and points of view and at different levels of abstraction and detail, to accommodate access to a specific group of collections. In any wider search accessing distributed collections, the user would like to start with familiar terminology and let the system find out the correspondences to other terminologies in order to retrieve equivalent results from all addressed collections. Hjørland showed in his article (2002:425) the relationship between specific domain vocabularies exemplified by the thesaurus and his theory of Domain Analysis: "thesauri are mainly domain specific vocabularies, and the methodology of designing them can be seen as one (implicit) form of domain analysis."

In Mustafa El Hadi (2008: 303) I mentioned the importance of reference tools, such as terminologies and thesauri for specialized information and communication. I wrote: "One of the purposes of information and communication sciences is the construction and sharing of meaning. This process involves building reference tools such as thesauri for indexing and retrieving an increasing digital data. Meaning construction through these tools is based on well known conceptual methods. To

guarantee communication between the various actors of a field of knowledge, or what we can call a ‘community of experts’ which uses its own discourse these tools should be rigorously and coherently designed”. I defined in the same paper (2008: 302) “discourse community” as “ a concept describing all forms of communication that contribute to a particular, institutionalized way of thinking; and ‘community’, which in this case refers to the people who use, and therefore help to create a particular discourse. (…) If one wants to become a member of a certain discourse community, it requires more than learning the lingo. It requires understanding concepts and expectations set up within that community”. This point of view had been described by Smiragilia (2012: 116) as a domain analytical as pertaining to Domain Analytical approach in knowledge organization.

8. Conclusion

Semantic interoperability has a broad scope and we consider it as one of the core elements towards cultural interoperability. It should be noted however, that full semantic interoperability cannot operate in all contexts in spite of the efforts made in that direction and that there is even more to say about the difficulties ahead in trying to achieve “cultural interoperability”. What hinders semantic interoperability and consequently cultural interoperability is that the degree of success achievable, in the integration of multiple knowledge representation systems or knowledge organization schemes, is constrained by limitations on the universality of human conceptual systems. We believe that combining the two types of KOS (classifications & language-based KOS such as multilingual thesauri and vocabularies) could progress towards full interoperability. Web ontology standards, such as SKOS (Simple Knowledge Organization System) is designed specifically with the intention to provide a common data model for sharing and linking knowledge organization systems via the Semantic Web. Standards such as SKOS are also meant to be used as a vehicle for publishing and sharing knowledge organization systems that were not born digital. The considerations sketched above can open a path of research that until now has been largely focused on technical and semantic interoperability. However there is a lack of general theoretical framework for considering these developments. We suggest that cultural interoperability and its framework should become a new epistemological perspective in KO.

Acknowledgements: The author is deeply grateful for the interesting discussions engaged during the DIGIKO (DIGITAL Culture and KOS) project writing. Special thanks to Sylvie Davis, Laurence Favier, Andrea Scharnhorst, Philip Mayer, Fabrice Papy Aida Slavic, and Peter Stockinger of the DIGIKO project consortium. Many ideas in this paper also originate from our discussions in writing the call for papers for the Special issue on “Cultural Interoperability”, *Communication*, (University of Laval) with the co-editors, Laurence Favier and Dominique Vinck (forthcoming). The author thanks also Gérard Prunier, a political analyst and an independent international consultant on geopolitical affairs, for the fruitful discussions engaged about cultural interoperability in diplomatic, political and military fields.

References

Arms, W. Y., Hillmann, D., Lagoze, C., Kraft, D., Marisa, R., Saylor, J., et al. (2002). A spectrum of interoperability. *D-Lib Magazine, 8*(1), http://www.dlib.org/dlib/january02/arms/01arms.html.

Brodeur, J, Edwards, G., Bédard, Y., Bernard Moulin, B. (2003), Revisiting the Concept of Geospatial Data Interoperability within the Scope of Human Communication Processes, In *Transactions in GIS,* Blackwell Publishing, 2003, 7(2): 243–265

Chung, E., Moens, W. (2007), “The Semantics of Semantic Interoperability: A Two-Dimensional Approach for Investigating Issues of Semantic Interoperability in Digital Libraries, Proceedings of the American Society for Information Science and Technology, http://www.unt.edu/wmoen/publications/Chung_Moen_ASIST_Short_Final.pdf

Beghtol, B. (2002a),"A proposed ethical warrant for global knowledge representation and organization systems", *Journal of Documentation*, Vol. 58 Iss 5 pp. 507 – 532

Beghtol, B. (2002b), “Universal Concepts, Cultural Warrant and Cultural Hospitality” in *Advances in KO*, Proceedings of the Seventh International ISKO Conference 10-13 July 2002 Granada, Spain Organized by the ISKO - Spanish Chapter and The University of Granada, p.45-49

Broughton, V. (2010), Emergent Controlled Vocabulary, in *LCN*, (Mustafa El Hadi & Hudon (ed), Vol. 6-N°3/2010 :50)

Carney, D.R., Hall, J.A., LeBeau, L. (2005). Beliefs about the nonverbal expression of social power. *Journal of Nonverbal Behavior*, 29, 105–123.

Coates, E.J. (1988),"The role of classifications in information retrieval : Action and thought in the contribution of Brian Vickery, *Journal of Documentation*, Vol. 44 Iss 3 pp. 216 – 225 Permanent link to this document: http://dx.doi.org/10.1108/eb026826 Downloaded on: 03 August 2015

Dagobert Soergel, Boris Lauser, Anita Liang, Frehiwot Fisseha, Johannes Keizer and Stephen Katz Reengineering Thesauri for New Applications: the AGROVOC Example available at http://www.fao.org/docrep/008/af234e/af234e09.htm#TopOfPage

Dextre Clarke, S. (2011) "ISO 25964: a standard in support of KOS interoperability", in A. Gilchrist and J. Vernau (eds), *Facets of Knowledge Organization, 4-5 July 2011, London*. London, Emerald. http://www.iskouk.org/conf2011/papers/dextreclarke.pdf.

Dextre Clarke, S., Zeng, M.L. (2012), "From ISO 2788 to ISO 25964: The Evolution of Thesaurus Standards towards Interoperability and Data Modeling", in *Information Standards Quality*, Special Edition: Year in Review and State of the Art of the Standards, p.20-26.

Doueihi, M. (2011). *Pour un humanisme numérique*. Paris: Seuil.

Doerr, M. (2001). Semantic problems of thesaurus mapping. Journal of Digital, Information, 1(8). http://jodi.tamu.edu/Articles/v01/i08/Doerr/.

DIGIKO (2013) Pré-Projet DIGIKO, un projet de recherche (Digital Cultures and Universality in KO) proposé avec quatre partenaires (la France en tant que pilote, le Royaume-Uni, les Pays Bas et l'Allemagne, conjointement avec un panel d'experts internationaux. Une version actualisée a été présentée durant la conférence *Knowescape,* à La Hague, le 23 octobre 2013.

Favier, L. and Mustafa El Hadi, W. (2013) "L'interopérabilité des systèmes d'organisation des connaissances : une nouvelle conception de l'universalité du savoir ?", in F. Papy (ed), *Problématiques émergentes dans les sciences de l'information*. Paris, Hermès.
Fox, E.A. Marchionini, G. (1998), " Toward a Worldwide Digital Library, in Communication of the" ACM April 1998/Vol. 41, No. 4, p.29-32

Friesen, N 2002. Semantic Interoperability and Communities of Practice. Global Summit Conference, Adelaide.http://www.educationau.edu.au/globalsummit/papers/nfriesen.htm

Green, R. (2002), Conceptual Universals in Knowledge Organization and Representation, in *Advances in Knowledge Organization*, in *Advances in KO*, Proceedings of the Seventh International ISKO Conference 10-13 July 2002 Granada, Spain Organized by the ISKO - Spanish Chapter and The University of Granada, p.15-27.

Green, R., Bean, C.A., Hudon, M. (2002) Universality and Basic Level Concepts (2002), in *Advances in KO*, Proceedings of the Seventh International ISKO Conference 10-13 July 2002 Granada, Spain Organized by the ISKO - Spanish Chapter and The University of Granada, p.311-317.

Hjørland, B. (2003) Fundamentals of Knowledge Organization, *Knowl. Org*. 30(2003)No.2

Hjørland. B. (2008) What is Knowledge Organization (KO)? *Knowl. Org*. 35(2008)No.2/No.3

Hjørland, B. (2002),"Domain analysis in information science: Eleven approaches –traditional as well as innovative", *Journal of Documentation*, Vol. 58 Iss 4 p422-462

Hodge G., (2000), "Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files", *The Digital Library Federation Council on Library and Information Resources* http://www.clir.org/pubs/abstract/pub91abst.html

Hoolande, S.V. De Wilde, Verborgh, Mannens, De Walle, & Hercher, (2013) “Evaluating the Success of Vocabulary Reconciliation for Cultural Heritage Collections”, *JASIST*, 64(3):464–479, 2013

Hudon, M. (1997). Multilingual thesaurus construction: Integrating the views of different cultures in one gateway to knowledge concepts. *Knowledge Organization*, 24, 84–91.

Hudon, M. (2005). Accès thématique en bibliothèque numérique : le rôle du langage documentaire de type « thésaurus ». *Les Bibliothèques numériques*, sous la dir. de Fabrice Papy, 151-177. Paris : Hermès Science.

Hudon, M. (2012), ISO 25964 : pour le développement, la gestion et l’interopérabilité des langages documentaires, in : *Documentation et bibliothèques*, n° thématique Normalisation et normes pour la gestion de l’information et des documents, vol. 58/3, 2012, 130-140.

Immler, M. (1991). “Is semantics universal, or isn’t it? On the relation of language, thought and semantic structure”. In D. Zaefferer (Ed,), *Semantic Universals and Universal Semantics*, 37-59. Berlin, New York

Dahlberg, I. Siwek, K; (editors) (1995), *Compatibility and Integration of Order Systems,* Research Seminar Proceedings of the TIP/ISKO Meeting, Warsaw, 13-15 September 1995, organized by The Polish Library, the Society for Professional Information Association.

Miller, P 2002, Interoperability. What is it and Why should I want it? *Ariadne* Issue 24. http://www.ariadne.ac.uk/issue24/interoperability

Koussouris, S. Lampathaki, F., Mouzakitis, S., Charalabidis, Y., Psarraa, J., (2011), “Digging Into The Real-Life Enterprise Interoperability Areas Definition And Overview Of The Main Research Areas” available at https://www.academia.edu/2852773/

Mustafa El Hadi (2008), “Discourse Community Analysis Sense Construction Versus Non-Sense Construction”, in *Advances in Knowledge Organization*, Vol. 11 (2008) “Culture and Identity in Knowledge Organization”, (Tennis & Arsenault editors), Proceedings of the Tenth International ISKO Conference, 5-8 August 2008 Montréal, Canada, p.302-306.

Mustafa El Hadi, W. (2013), “Faceted Classifications as a Model for Universal Classifications”. In: From Libraries to the Web, *SRELS Journal of Information Management*, Vol. 50, (6); December 2013; p721-734.*:*

Papy, F. (2013), “Classification and visualization: augmenting user independence and enhancing collections use”, in Proceedings of UDCC Seminar, *Classification & Interfaces to Knowledge,* The Hague, 23-24 October 2013.

Pedro, R.D. (1999) "The Translatability of Texts: A Historical Overview", Meta : journal des traducteurs / Meta: Translators' Journal, vol. 44, n° 4, 1999, p. 546-559.

Slavic, A. (2011) “Classification revisited: a web of knowledge”, in A. Foster and P. Rafferty (eds), *Innovations in information retrieval: perspectives for theory and practice*. London, Facet. pp. 23-48.

Slavic, A. (2008). Faceted Classification: management and use. *Axhiomathes*, vol. 18, p. 257-271.

Slavic, A. (2011). Classification revisited: a web of knowledge. In: *Innovations in information retrieval: perspectives for theory and practice* . Ed. by A. Foster and P. Rafferty P. London: Facet. pp. 23-48.

Slavic, A (2013) Classification in knowledge transfer between languages, domains and collections in humanities: from legacy collections to the Web of linked data, Presentation at European Seminar, Ecole Doctorale, 12th April, 2013, Lille

Smiragilia, R. (2012) “Epistemology of Domain Analysis”. in *Cultural Frames of Knowledge”,* Smiragilia & Lee editors), Ergon Verlag, pp 111-124

Soergel, D. (2010) “Concepts in context: proceedings of the Cologne Conference on Interoperability and Semantics in Knowledge Organization, July 19th-20th, 2010, Ergon Verlag Publisher, 13-36
Tennant, R. (2001). Different paths to interoperability. *Library Journal, 126*(3), 118-119.

Tennis, J. T (2003): Two Axes of Domains for Domain Analysis, Knowl. Org. 30(2003) No.3/No.4 191-195.

Vizine-Goetz, D., Hickey, C., Houghton, A., Roger, Th. (2004) “Vocabulary Mapping for Terminology Services”, *Journal of Digital Information*, 4 (4).

Wimmer, A. (2008), “Elementary strategies of ethnic boundary making”, in: *Ethnic and Racial Studies*, 31:6, 1025 —1055.

Winslow, D. and Everts, P. (2001) “It’s not a Question of Muscle: Cultural Interoperability for NATO”. In NATO - The First Fifty Years. G. Schmidt, ed., Hampshire, UK: Palgrave/MacMillan pp. 85-104.

Zeng, M.L., Chan, L.M., (2004) “Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems, *JASIST*, 55(5):377–395, 2004.

Reflexões acerca do desenvolvimento de metodologia para análise de assunto

Paula Regina Dal'Evedove
Universidade Federal de Pernambuco
paula.dalevedove@ufpe.br

Roberta Cristina Dal'Evedove Tartarotti
Universidade Estadual Paulista
roberta_tartarotti@yahoo.com.br

Mariângela Spotti Lopes Fujita
Universidade Estadual Paulista
fujita@marilia.unesp.br

1 Introdução

As investigações em torno do indexador como leitor profissional oferecem observações importantes a respeito da mente humana, bem como considerações para o aprimoramento dos procedimentos da análise de assunto. Diante disto, um dos grandes desafios para a área de Organização e Representação do Conhecimento enquanto disciplina científica e social de reflexão e produção teórica (GARCÍA MARCO, 1995), confluente ao entendimento das problemáticas que envolvem o conhecimento registrado, centra-se na criação de metodologias destinadas ao processo de indexação que fomentem o processamento intelectual da análise de assunto.

O cenário teórico e metodológico da Indexação, enquanto área de estudo e investigações, apresenta orientações que viabilizam os procedimentos de análise,

síntese e representação do conteúdo informacional dos documentos na prática profissional do bibliotecário indexador. Apesar dos avanços, algumas questões permanecem abertas no campo teórico e normativo deste processo de organização temática da informação, como é o caso da subjetividade na compreensão de conceitos que por tempos está presente nas discussões da área e acompanha o indexador ao longo de sua trajetória profissional. A notoriedade da subjetividade no processo de indexação decorre de sua concatenação com diversos outros aspectos de cunho social, cultural, político, físico, linguístico, lógico e cognitivo, aplicáveis à compreensão de assuntos e que condicionam o resultado da indexação.

Envolvida por suas dimensões teóricas e aplicáveis, a análise de assunto na prática de indexação decorre de processos de compreensão permeados por atos subjetivos procedentes do conhecimento prévio e da formação e experiência profissional, bem como do contexto cultural e ideológico do bibliotecário indexador e do sistema de informação. Deste modo, o desenvolvimento de metodologias destinadas à análise de assunto de efetivo controle dos elementos que permeiam à compreensão de conceitos é um desafio a ser encarado na Organização e Representação do Conhecimento. Fujita (2013, p. 156) defende que a natureza do processo de indexação, ainda que "[...] empírico e sujeito a uma aplicabilidade intuitiva demanda investigações científicas que se proponham a aprimorar o conhecimento profissional de quem o executa, bem como os instrumentos, métodos e técnicas utilizados para tal". Contudo, a partir de quais aspectos teóricos e metodológicos é possível circunscrever as ações subjetivas dos bibliotecários indexadores que atuam em contextos de informação permeados por aspectos socioculturais?

Em virtude da relevância coberta pelo tema, trabalha-se com o repertório da análise de assunto no que tange à compreensão de conceitos, buscando tecer considerações sobre as referências socioculturais presentes nas ações profissionais. A intenção neste estudo não é dar respostas ao problema suscitado, mas refletir acerca dos desafios que cercam a criação de instrumentos destinados à análise de assunto, diante da diversidade cultural com a qual o bibliotecário indexador lida na prática de indexação no âmbito de domínios informacionais contemporâneos.

2 A compreensão de conceitos na prática de indexação

> A leitura documental faz parte das atividades desempenhadas pelo bibliotecário indexador, sendo considerada a fase inicial que desencadeia todas as outras etapas do processo de indexação e que envolve três aspectos principais: a compreensão do conteúdo do documento, a identificação dos conceitos que representem este conteúdo e a seleção dos conceitos válidos para a recuperação em sistemas de informação –, superpostos durante a leitura do documento (FUJITA, 2003b, p. 64; TARTAROTTI, 2014, p.147).

Giasson (1993) defende a visão interacionista da leitura enquanto um processo de comunicação e interação entre três variáveis interligadas: texto-leitor-contexto. Na análise de assunto, quanto maior a interação destas variáveis durante a leitura documental, maior o nível de compreensão profissional acerca do assunto do documento. De modo geral, a variável texto representa a estrutura textual na leitura documental e contém as ideias do autor, em que o conhecimento textual faz parte do conhecimento prévio do leitor e propicia condições de compreensão quando o mesmo possui habilidade e familiaridade com as diferentes tipologias e estruturas textuais, o que facilita a compreensão do assunto. Entretanto, “a recepção da informação obtida do texto através da leitura dá-se de forma diferenciada de leitor para leitor, pela maneira de cada um atribuir sentido a seu conteúdo” (DIAS, NAVES, 2007, p. 47).

A interação do bibliotecário indexador com o texto enquanto um leitor profissional ocorre mediante o uso de estratégias metacognitivas e de seu conhecimento prévio, mantendo em mente o seu objetivo de representar o texto para futura recuperação e, também, considerando as limitações da atividade e os objetivos do domínio informacional no qual está inserido (FUJITA, 2003a). Esta interação consiste numa amplitude de procedimentos coordenados e de operações perceptíveis, linguísticas e cognitivas, em que as características da estrutura textual fornecem subsídios ao conhecimento do bibliotecário indexador, aumentando a compreensão do assunto abordado.

A compreensão não depende apenas das características intrínsecas do texto, mas do conhecimento prévio compartilhado entre o autor e o leitor, em que este possui em sua concepção uma representação de mundo e compreender um texto é relacionar elementos dessa representação. A leitura é uma prática que visa dar sentido perceptivo e intelectual ao texto. Neste contexto insere-se o leitor e sua capacidade subjetiva de interpretar. Desse modo, o leitor é a segunda variável do processo de interação munido de objetivos para a leitura documental, sendo o elemento de maior influência e capaz de interferir na representação documental.

Por sua vez, a variável contexto refere-se à análise de assunto em contextos de informação especializados e todos os elementos influentes no momento da leitura profissional, isto é, “[...] todas as condições nas quais se encontra o leitor quando entra em contato com um texto” (GIASSON, 1993, p. 40). Na concepção de Van Dijk (1997), é o conjunto das propriedades da situação social, sendo estas relevantes para a produção, compreensão e/ou funcionamento do discurso e de suas estruturas. Diante disso, o contexto representa não apenas o conhecimento prévio profissional e os objetivos da atividade em questão, mas o ambiente em que é realizada esta atividade profissional em leitura documental.

Para que o significado do conteúdo informacional de um documento seja o menos ambíguo possível, o mesmo deve ser analisado de acordo com o contexto no qual está inserido, uma vez que tal variável tende a agregar valor em determinados conceitos, os quais passam a assumir novos significados apenas mudando-se o ambiente de análise (PINTO MOLINA, 1993). Portanto, o conceito somente assume sentido quando levado em consideração o contexto de situação, em que o bibliotecário indexador deve ajustá-lo as suas decisões conforme a necessidade exigida pelo meio.

Com efeito, as etapas de identificação e seleção de conceitos devem ser dirigidas para o processo de compreensão e investigadas sob a ótica de outros campos científicos, em que o estudo da leitura documental torna-se interdisciplinar quando considerado uma prática de dar sentido perceptivo e intelectual ao texto pelas ações do bibliotecário indexador e sua capacidade subjetiva de interpretar, bem como das especificidades das tipologias textuais e do domínio informacional em que a análise de assunto ocorra. O conhecimento prévio/cognição profissional faz parte do contexto sóciocognitivo enquanto uma representação mental do leitor. Sendo assim, deve-se considerar a visão teórica de Hjørland (2002) quanto à proposta de interação entre o sujeito e o ambiente social por dois aspectos: a prática de indexação é um processo intelectual que depende da cognição e; o domínio do contexto contribui para que a identificação e seleção de conceitos abarquem as necessidades dos usuários e do ambiente de enunciação.

3 Desafios para a construção de metodologia para análise de assunto

Existem iniciativas teóricas e aplicáveis importantes que suscitam reflexões e lançam luz à concretização de uma metodologia para análise de assunto que cerque os principais aspectos que permeiam a prática de indexação. Contudo, as pesquisas em análise de assunto tendem a verificar os padrões de processamento do indexador (DIAS; NAVES, 2007), isto é, busca-se identificar as estratégias capazes de extrair de um documento o seu assunto – foco no processo.

Apesar dos avanços na literatura especializada, compreende-se muito pouco sobre as ações do indexador e sua influência no resultado final do processo de indexação, em especial quando se admitem os problemas da organização do conhecimento na contemporaneidade, muitos destes ligados às "[...] dificuldades de tratar documentos em face à manifestação das diversidades sociais, culturais e políticas" (LARA, 2013, p. 238). Portanto, ainda são incipientes as pesquisas sobre a etapa da análise de assunto que cerquem as questões sobre a ação humana a partir da realidade social objetiva. Dal'Evedove e Fujita (2013, p. 44) compreendem que esta "[...] mudança paradigmática é necessária para a compreensão da realidade profissional em domínios informacionais específicos", considerando-se que aprimoramentos e melhores especificações no desenvolvimento da prática de indexação são necessidades emergentes que merecem um olhar mais atento da área.

Neste sentido, torna-se importante o encaminhamento de pesquisas na área de Organização e Representação do Conhecimento que busquem mensurar o verdadeiro impacto das ações particulares do bibliotecário indexador na prática de indexação, atrelando as características da abordagem sociocognitiva com os aspectos de natureza sociocultural que cercam o contexto informacional. Isto porque, entende-se que a análise de assunto não pode ser estudada separadamente do contexto social em que a indexação é realizada. Com efeito, além das questões relativas à compreensão de conceitos na prática de indexação e que estão diretamente relacionadas ao indexador na qualidade de leitor profissional, as pesquisas destinadas ao tema precisam pensar em novas possibilidades investigativas que permitam reconhecer as necessidades informacionais dos usuários frente aos diferentes domínios informacionais.

A este respeito, Carvalho (2004, p. 165) advoga que "[...] fica contraditório oferecer serviços e produtos para atender às necessidades informacionais de usuários que não são efetivamente conhecidos, nem em suas características minimalistas".

Porém, apesar de demandar esforços, reconhecer o usuário como elemento norteador da prática de indexação é a melhor alternativa para suprir deficiências e melhorar a qualidade dos produtos informacionais gerados, com vistas à ampliação do uso. Complementando a questão, a referida autora sinaliza que a oferta de produtos e serviços prontos e, sobretudo, "[...] testados na perspectiva somente do bibliotecário, como algo imutável ou passível apenas de pequenas adaptações" precisa ser abandonada nas práticas biblioteconômicas.

Observa-se que a análise de assunto está condicionada a determinados objetivos, contextos organizacionais e ao sistema de recuperação da informação e, portanto, na prática de indexação as questões sociocognitivas e socioculturais se entrelaçam. Neste sentido, para o enfrentamento de algumas lacunas de ordem teórica e normativa que permeiam a indexação de assunto, considera-se relevante o encaminhamento de pesquisas pela abordagem sociocultural, especificamente quando o objetivo é a criação de metodologias para análise de assunto. Isto porque, "o valor da informação é dependente do significado atribuído pelo sujeito, o que implica conceber a informação em seu contexto de enunciação para vinculá-la à cultura específica de cada domínio do conhecimento" (DAL'EVEDOVE, 2014, p. 116).

Dessa forma, versar as problemáticas de ordem sociocultural na análise de assunto é um caminho salutar para a área, em razão da necessidade de se refletir sobre a atuação do bibliotecário indexador diante de uma atividade profissional exercida em domínios informacionais diversos. Admite-se, portanto, que o viés sociocultural aplicado aos estudos da análise de assunto é uma importante alternativa para enfrentar as configurações contemporâneas da área de Organização e Representação do Conhecimento, mediante a possibilidade de se conhecer os modos sociais de pensamento dos bibliotecários indexadores e dos usuários; atores sociais que apresentam opiniões, impressões, comportamentos e interesses pessoais e coletivos diversos.

4 Apontamentos finais

Diante das reflexões apresentadas, aceitar e reconhecer que a atividade de indexação abriga questões de natureza sociocognitiva e sociocultural é um caminho necessário às pesquisas dedicadas à temática. Investigar a prática de indexação sob

o viés da abordagem sociocultural é viabilizar que, além das problemáticas acerca do processo, questões relativas aos atores sociais e ao domínio informacional sejam efetivamente consideradas, visando contribuir com subsídios teórico-metodológicos que busquem suprir algumas das deficiências que permanecem em aberto no âmbito da Indexação.

A observância de aspectos cognitivos, culturais e sociais que permeiam a prática de indexação, relativizados, é uma forma de se considerar as relações que os atores sociais estabelecem com os conteúdos informativos presentes nos documentos. Neste sentido, entende-se que para o desenvolvimento de uma metodologia destinada à análise de assunto, os aspectos referenciados precisam ser considerados.

Ademais, o aperfeiçoamento da prática de indexação está diretamente ligado ao desenvolvimento de metodologias destinadas à etapa da análise de assunto, pois a partir dela o profissional identifica e seleciona conceitos representativos do documento, os quais serão traduzidos em termos descritores por meio de uma linguagem de indexação. Logo, pesquisas na temática que sejam encaminhadas sob o viés sociocultural estarão em consonância com o preconizado por Saracevic (1999), para o qual os estudos empreendidos em Ciência da Informação devem integrar as várias manifestações e comportamentos do fenômeno informação (análise da informação) e, paralelamente, reconhecer os comportamentos, efeitos e as interfaces dentro de um contexto de dimensões sociais.

Sendo assim, a presente pesquisa abarca questões-chave presentes no âmbito da Indexação de dimensão teórico-conceitual e prático-aplicada, entrelaçadas por mudanças contínuas nos domínios informacionais específicos em meio à composição e recomposição da sociedade na contemporaneidade, discussões estas ainda dispersas, com entendimentos a consolidar.

Referências

CARVALHO, I. C. L. *A socialização do conhecimento no espaço das bibliotecas universitárias.* Rio de Janeiro: Interciência, 2004. 185 p.

DAL' EVEDOVE, P. R. *O tratamento temático da informação em abordagem sociocultural:* diretrizes para definição de política de indexação em bibliotecas universitárias. 2014. 266f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

DAL' EVEDOVE, P. R; FUJITA, M. S. L. Estudo sociocultural da comunidade discursiva do tratamento temático da informação em bibliotecas universitárias. *Encontros Bibli:* revista eletrônica de biblioteconomia e ciência da informação, v. 18, n. 36, p. 23-50, jan./abr., 2013.

DIAS, E. W.; NAVES, M. M. L. *Análise de Assunto:* teoria e prática. Brasília: Thesaurus, 2007. 116 p.

FUJITA, M. S. L. A importância teórica e prática da indexação na fundamentação científica da organização e representação do conhecimento. In: DODEBEI, V.; GUIMARÃES, J. A. C. (Org.). *Complexidade e organização do conhecimento:* desafios de nosso século. Rio de Janeiro: ISKO-Brasil; Marília: FUNDEPE, 2013. p. 147-159.

FUJITA, M. S. L. *A leitura documentária do indexador:* aspectos cognitivos e linguísticos influentes na formação do leitor profissional. 2003. 321 f. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabéticas) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003a.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para a indexação. *Revista Digital de Biblioteconomia e Ciência da Informação*, Campinas, v. 1, n.1, p. 60-90, 2003b.

GARCIA MARCO, F. J. Los contenidos y la secuencia de la organización del conocimiento: una propuesta interdisciplinar. In: GARCIA MARCO, F. J. (Ed.). *Organización del conocimiento en sistemas de información y documentación*. Zaragoza: Librería General, 1995. p. 219- 228.

GIASSON, J. *A compreensão na leitura.* Lisboa: Asa, 1993. 317 p.

HJØRLAND, B. Epistemology and the socio-cognitive perspective in information science. *Journal of the American Society for Information Science and Tecnology*, v. 53, n. 4, p. 257-270, 2002.

LARA, M. G. de. Problemas da organização do conhecimento na contemporaneidade. In: DODEBEI, J. A. C. G. (Org.). *Complexidade e organização do conhecimento:* desafios de nosso século. Rio de Janeiro: ISKO-Brasil; Marília: FUNDEPE, 2013. p. 238-245.

PINTO MOLINA, M. *Análisis documental:* fundamentos y procedimientos. 2. ed. rev. y aum. Madrid: EUDEMA, 1993. 270 p.

SARACEVIC, T. Information science. *Journal of the American Society for Information Science*, v. 50, n. 12, p.1051-1063, 1999.

VAN DIJK, T. A. (Ed.). *Discourse as structure and process.* London: Sage, 1997. 368 p.

Organização, representação, recuperação e acesso à informação: (re)configuração do Formato MARC21 e do BIBFRAME pelos propósitos da diversidade cultural nos ambientes informacionais digitais?

Zaira Regina Zafalon
Universidade Federal de São Carlos
zaira@ufscar.br

Marcela Cristina Néspoli
Universidade Federal de São Carlos
marcela_nespoli@hotmail.com

1 Introdução

Estudos de organização e representação da informação tem sido ressignificados ao longo dos tempos em busca de garantia da recuperação e do acesso à informação principalmente em decorrência da diversidade presente nas instituições de patrimônio cultural. Tal diversidade pode ser caracterizada pelo comportamento de busca, pelo modo de recuperação e acesso à informação, pela (re)configuração da produção informacional e pelas tecnologias adotadas. Desta forma, conceitos, modelos, padrões e formatos de organização e representação da informação também têm sido discutidos. Esta pesquisa insere-se neste universo e busca discutir os propósitos do Formato MARC21 e do BIBFRAME enquanto padrões de representação da informação em ambientes informacionais digitais. As discussões envolvem, como pano de fundo, a ampliação do universo de busca e recuperação da informação pelas comunidades bem como o uso e reuso de metadados pelos parques tecnológicos; ambos cingem relação com a disponibilização da informação pelas unidades de patrimônio cultural na web por meio da descrição dos recursos. Para o atendimento ao objetivo proposto a pesquisa apresenta abordagem qualitativa, de natureza aplicada, com perfil descritivo e exploratório. Recorre, a um só tempo, à pesquisa

bibliográfica e documental. Por ser um tema ainda recente o presente trabalho pretende servir de contribuição social e teórica para que futuras pesquisas possam ser realizadas acerca deste tema.

2 Antecedentes sobre representação de recursos

Ações de organização e representação da informação (ORI) pautam-se em processos de análise, síntese, condensação, representação, recuperação, uso e reuso de metadados dos recursos informacionais em prol da convivência do usuário e da obtenção de documentos que busquem atender suas necessidades informacionais dadas as possibilidades de estabelecimento de comunicação. Destacam-se os estudos voltados aos conceitos, modelos, padrões e formatos de organização e representação da informação, quer sejam aqueles voltados à leitura, análise, condensação, representação e construção das linguagens documentárias, quer sejam aqueles dos requisitos funcionais, da modelagem conceitual, da interoperabilidade e das estruturas de compartilhamento e acesso à informação por humanos e máquinas.

O resultado dos esforços da representação documental, presente nos estudos de ORI, concentra-se no registro bibliográfico que estabelece, portanto, *relações semânticas*, entre os elementos presentes no documento representado e o seu valor, *subsemânticas*, marcadas pela relação entre as várias unidades que compõem um registro bibliográfico, e *supersemânticas*, entre os vários registros que se referem às várias manifestações de uma obra, além das *relações sintáticas*, requisito para o registro do valor nas estruturas semânticas. (ZAFALON, 2014).

Os padrões de representação, tanto os de conteúdo quanto os de estrutura, garantem a consistência dos registros, definindo regras de sintaxe e semântica para a descrição do suporte informacional e do conteúdo dos recursos informacionais, e atestam o sucesso na recuperação de recursos, uma vez que minimiza eventuais diferenças na representação e no compartilhamento da informação.

3 (Re)configuração de propósitos pela diversidade cultural?

O Formato MARC21, enquanto padrão de estrutura de metadados, tomado como requisito para o processo de compartilhamento de registros bibliográficos, provê mecanismos de leitura e interpretação de dados bibliográficos por máquinas. Aliado ao formato de intercâmbio de registros (ISO 2709), aos padrões de conteúdo (CDD, CDU, AACR, LCSH, dentre outros) e aos protocolos de comunicação e recuperação bibliográfica (Z39.50), garante a transferência de dados qualificados entre máquinas. Avram (1970, 1975) destaca que o desenvolvimento do Formato baseou-se na filosofia de compartilhamento e reuso da informação, na qual um conjunto de dados pode ser acessado e utilizado por vários usuários ao mesmo tempo. Inicialmente configurado para uso pela Library of Congress (LC), o MARC assume características que permitem que cada biblioteca tenha autonomia tanto no provimento de registros quanto no compartilhamento com outras unidades. A flexibilidade do MARC, segundo Zafalon (2008), se dá por meio de suas marcações, haja vista as convenções adotadas que tornam o gerenciamento e a recuperação dos dados bibliográficos mais eficientes, resultante da grande especificidade que promovem. O MARC foi, portanto, desenvolvido com o intuito de padronizar a representação dos registros informacionais e, assim, permitir, por meio de ações cooperativas, o intercâmbio eficiente de dados bibliográficos de modo a evitar a duplicidade de esforços. Desvantagens do Formato MARC, discutidas por Tennant (2002, 2013), um dos apoiadores da iniciativa para substituição do formato MARC, indicam a complexidade de sua adoção, a dependência das marcações para efeitos semânticos, a manipulação dos registros feita por profissionais qualificados, o que gera deserção e resistência da própria comunidade, e a dificuldade de leitura do registro dada a sintaxe complexa, natural do esquema numérico dos rótulos. Destaca-se também a incapacidade de o formato realizar relacionamentos de documentos que são inteiramente digitais, como as URLs.

Neste contexto, discute-se como substituto ao MARC o modelo Bibliographic Framework (BIBFRAME), iniciativa da LC, desde 2012, cujo objetivo é o de servir como modelo para o qual vários modelos de conteúdo possam ser mapeados. Está sendo projetado para, intencionalmente, tornar-se ponto de entrada formal da comunidade bibliotecária em uma rede que envolve tanto metadados quanto documentos lincados. Isso tende a se tornar possível pelo fato de o BIBFRAME basear-se na proposta de criação de relações semânticas (LinkedData) e por ter modelagem baseada no Resource Description Framework (RDF), modelo para

intercâmbio de dados na web. Diante do paradigma de reuso da informação em seu mais alto nível, o BIBFRAME reconhece entidade, atributos e relacionamentos, por meio de suas classes principais (obra, instância, autor e anotação). (LIBRARY OF CONGRESS, 2012).

Ao alinhar-se o objetivo do MARC à necessidade de disponibilizar os recursos bibliográficos aos usuários, à possibilidade de acesso aos recursos bibliográficos e, por fim, ao objetivo de garantir acesso aos recursos bibliográficos, entende-se que o MARC atende ao seu propósito inicial, uma vez que garante a recuperação e o acesso aos recursos bibliográficos nele descritos. Todavia, em um contexto mais abrangente que envolve estudos de *open data* e *web de dados*, p. ex., o MARC aparece como obsoleto dada a dificuldade de criar relacionamentos entre os recursos, o que compromete o objetivo de navegação para além do catálogo. O BIBFRAME, contudo, em termos de estruturação dos dados bibliográficos, com vistas a permitir o intercâmbio de dados, bem como possibilitar a recuperação de informações cada vez mais específicas, indica, por enquanto uma lacuna no novo modelo, o que tem gerado debates e discordâncias.

4 Discussões finais

O Formato MARC tem sofrido duras críticas pela quantidade de *tags*, campos e subcampos nem sempre utilizados e que acarretam custos desnecessários aos softwares, além da dificuldade para os catalogadores em usá-lo. O BIBFRAME, por conta de prever pouca especificidade na descrição dos registros, compromete o resultado de busca e o refino de pesquisa, que requerem granularidade fina. O BIBFRAME prevê a descrição nas instâncias e nos relacionamentos com outras classes principais (obra, autor e anotação), mas não a descrição específica. Isso é decorrente da modelagem de dados utilizada, o RDF, o qual não comporta a criação de tantos predicados quantos forem necessários para alcançar a especificidade que se teria com o MARC. Outro ponto refere-se à temporalidade das URLs e a padronização na criação das URIs, afim de garantir a univocidade na recuperação, tal qual previsto no MARC. Obter controle de autoridades e de acesso é imprescindível para o alcance do objetivo do modelo.

Mesmo diante de tantos questionamentos e incertezas, a LC e a Zepheria (empresa especializada em aplicativos para bibliotecas) estão desenvolvendo uma versão *demo* do modelo para a descrição de recursos bibliográficos a fim de estruturar os dados e permitir a entrada dos mesmos. Não fica claro, porém, como se dará o armazenamento dos dados e o intercâmbio com outros formatos. A dificuldade estaria no fato de o RDF ter por função descrever relacionamentos genéricos entre os dados, permitindo que computadores utilizem e troquem esses dados e ainda preserve o seu significado original, mas não estrutura a informação.

Com o intuito de se traçar uma comparação entre o MARC21 e o BIBFRAME observa-se, quanto ao propósito, que o primeiro dedica-se a permitir a comunicação de dados bibliográficos por meio do intercâmbio de modo a atender a qualquer necessidade informacional, em qualquer lugar do mundo, por qualquer biblioteca ou usuário; e que o segundo busca integrar as várias comunidades do conhecimento, por meio dos dados lincados. Quanto à estrutura ambos diferenciam-se dado que o MARC se utiliza de campos e subcampos, enquanto o BIBFRAME adota as triplas RDF. As vantagens do MARC podem ser observadas diante da economia de tempo, inserção de múltiplos registros, consistência dos registros, garantia de integridade dos dados, possibilidade de catalogação cooperativa e garantia de recuperação mais eficiente. O BIBFRAME, por sua vez, apresenta integração de várias comunidades do conhecimento, possibilidade de proporcionar melhor navegabilidade, autossuficiência na descrição dos registros, entrada da biblioteca na Web de dados. O contraponto ao uso do BIBFRAME concentra-se na falta de familiaridade dos usuários com o RDF e o fato de este não ser estruturado para intercâmbio da informação, fazendo uso de pouca especificidade nas descrições.

Ao retomar-se o objetivo geral da pesquisa entende-se que o BIBFRAME surge com o intuito de integrar as bibliotecas a todas as comunidades do conhecimento, através da *web*, uma vez que, por conta de sua estrutura, o formato MARC21 não pode realizar tal feito. Deste modo, pode-se apurar que a proposta de substituição nasce baseada na necessidade de integração das bibliotecas com o ambiente *web* para poder permitir uma maior navegabilidade ao usuário, integrando conceitos e modelos tais como FRBR, ontologias, RDF, XML e LinkedData.

Nesse sentido, tanto MARC quanto o BIBFRAME se destinam ao propósito de integrar comunidades que se encontram ou se encontravam isoladas. O MARC, com

um propósito um tanto tímido se comparado ao BIBFRAME (ao desconsiderarmos a configuração cultural e tecnológica da época), conectou as bibliotecas para que pudessem intercambiar registros bibliográficos, o que acarretou em benefícios como consistência de dados, reuso das informações e diminuição do retrabalho. O BIBFRAME, por meio do conceito de navegabilidade, tende a integrar em um nível macro, a biblioteca às mais variadas comunidades do conhecimento. Contudo, a similaridade dos propósitos está restrita apenas a esse ponto, uma vez que ainda existem, no desenvolvimento do BIBFRAME, muitas lacunas e incertezas que não contemplam os objetivos atendidos pelo MARC e, portanto, não se designa ao mesmo papel desempenhado pelo formato.

Desse modo, não se vislumbra, ao menos por enquanto, que o BIBFRAME possa ser substituto do MARC. Entretanto, como também ocorreu com o MARC, muitas de suas aplicações e vantagens enquanto modelo de intercâmbio de dados não puderam ser sentidas até sua estabilidade e concepção como formato consolidado. O mesmo pode ocorrer com o BIBFRAME que, como modelo, não provê solução para todas as dúvidas que ainda pairam.

Em suma, embora a iniciativa BIBFRAME tenha surgido com a finalidade de atualizar o MARC21 entende-se que não haja impedimento da coexistência entre ambos, uma vez que cada um atende a propósitos específicos e complementares.

Referências

AVRAM, H. D. *MARC*: its history and implications. Washington: Library of Congress, 1975.

AVRAM, H. D. *The evolving MARC system*: the conception of a data utility. Illinois: University of Illinois, 1970.

LIBRARY OF CONGRESS. *Bibliographic Framework as a Web of Data*: Linked Data Model and Supporting Services. Washington, DC. 2012. Disponível em: <www.loc.gov/bibframe/pdf/marcld-report-11-21-2012.pdf>. Acesso em: 28 abr. 2015.

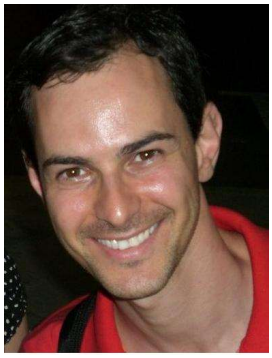
TENNANT, R. MARC Must Die. *Library Journal*, 15 Oct. 2002. Disponível em: < http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>. Acesso em: 01 out. 2014.

TENNANT, R. The Post MARC Era, Part 2: Where the problems lie, part 1. *Library Journal*, 8 May 2013. Disponível em: < http://www.thedigitalshift.com/2013/05/roy-tennant-digital-libraries/the-post-marc-era-part-2/>. Acesso em: 25 set. 2014.

ZAFALON, Z. R. *Formato MARC21 Bibliográfico*: estudos e aplicações para livros, folhetos, folhas impressas e manuscritos. São Carlos: EdUFSCar, 2008.

ZAFALON, Z. R. *Scan for MARC*: conversão de registros bibliográficos analógicos para o Formato MARC21 Bibliográfico. São Paulo: Editora Unesp, 2014.

Música, literatura e audiovisual: as contribuições da organização do conhecimento (OC) nas relações de interseccionalidade das obras de Dorival Caymmi e Jorge Amado



Fabio Assis Pinho
Universidade Federal de Pernambuco
fabiopinho@ufpe.br

Francisco Arrais Nascimento
Universidade Federal de Pernambuco
francisco.arrais.nascimento@gmail.com

Andréa Carla Melo Marinho
Universidade Federal do Rio Grande do Sul
andreacmmarinho@gmail.com

1 Introdução

O nordeste do Brasil fora evidenciado em sua construção historiográfica por suas características climáticas, econômicas e políticas, construindo assim uma imagem distorcida, alocando o mesmo como o espaço da seca, da pobreza e do coronelismo. Compreende-se que os ecos discursivos de tal estereotipagem influenciam a produção de artefatos culturais norteada pelos dispositivos de controle,

criando assim um universo de representações sociais da região. Segundo Albuquerque Jr. (2010, p.30):

> O discurso da estereotipia é um discurso assertivo, repetitivo, é uma fala arrogante, uma linguagem que leva à estabilidade acrítica, é fruto de uma voz segura e autossuficiente que se arroga o direito de dizer o que é outro em poucas palavras. O estereótipo nasce de uma caracterização grosseira e indiscriminada do grupo estranho, em que as multiplicidades e as diferenças individuais são apagadas, em nome de semelhanças superficiais do grupo.

Podemos compreender que os artefatos culturais representam não apenas formas de esteriotipagem, mas também elementos de resistência ao proselitismo social, marcos identitários e de representação social. Tal afirmação pode ser comprovada ao voltarmo-nos para os aspectos culturais do Nordeste do Brasil, onde atores sociais como Jorge Leal Amado de Faria (1912-2001) e Dorival Caymmi (1914-2008) fizeram de suas produções artísticas (literatura e música respectivamente), representações verossimilhantes do cotidiano baiano de onde emergem personagens marginalizados pelo sectarismo que delineava a estrutura social do Estado da Bahia no século XX, tais como retirantes, negros, coronéis, políticos, comerciantes, boêmios e prostitutas.

Portanto, o objetivo da pesquisa é analisar as relações de entre a produção de Jorge Amado e Dorival Caymmi em artefatos audiovisuais. Para isso, será utilizado um estudo de interseccionalidade estabelecido entre as produções dos artistas por meio das adaptações para o cinema e televisão da obra literária do primeiro que tiveram as canções do segundo como trilha sonora. Segundo CRENSHAW (2002, p.177):

> A interseccionalidade é uma conceituação do problema que busca capturar as conseqüências estruturais e dinâmicas da interação entre dois ou mais eixos da subordinação. Ela trata especificamente da forma pela qual o racismo,o patriarcalismo, a opressão de classe e outros sistemas discriminatórios criam desigualdades básicas que estruturam as posições relativas de mulheres, raças, etnias,classes e outras.

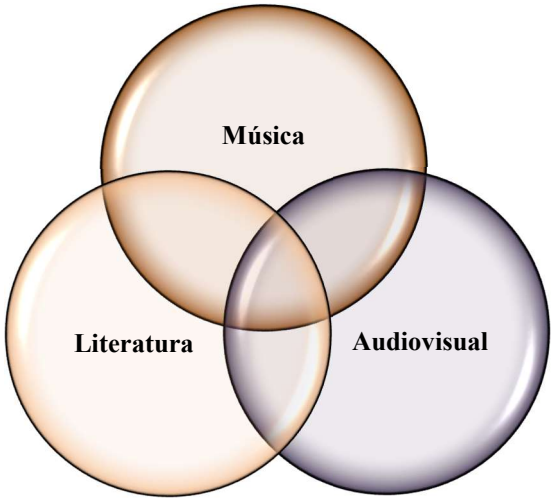


Figura 1 – Interseccionalidade entre artefatos culturais
Fonte: Elaborado pelos autores

Segundo Ferreira (2001) música é compreendida como arte e ciência de combinar os sons de modo agradável à audição, assim como escrita sob a forma de composição musical (música inscrita). O estudo aqui apresentado aponta a música sob a forma inscrita como recurso informacional, dado o seu caráter documental, enfocando assim a materialidade da música, compreendida sob a forma de inscrição do conhecimento (documento), seus elementos descritivos e temáticos analisados sob a óptica dos pressupostos teóricos e metodológicos da Organização da Informação (OI) e da Organização do Conhecimento (OC) para a compreensão e análise dos mesmos.

Segundo Ferreira (2001) a literatura é definida como sendo a arte de compor trabalhos artísticos em prosa e verso. Enquanto documento é caracterizada como obra literária e o tratamento temático se dá a partir da nacionalidade do autor e gênero literário.

Os artefatos culturais de natureza audiovisual têm sua gênese na produção fílmica, que segundo Moreno (1995) têm por objetivo transmitir uma mensagem através de uma composição de som e imagem, obedecendo a códigos que são estabelecidos pela técnica cinematográfica e tem origem nas demandas sociais.

2 Metodologia

No tocante aos procedimentos metodológicos recorreu-se às técnicas de cartografia de documentos e análise de assunto, entendendo a natureza descritiva do estudo. Portanto, trata-se de uma pesquisa de cunho documental, apoiada em pesquisa bibliográfica, que lança mão da combinação de metodologias que apresentam uma proposta de representação das estruturas documentais.

A partir do referido recorte metodológico, foram identificados 13 livros de Jorge Amado, 22 canções de Dorival Caymmi e 26 produções audiovisuais, das quais 15 são filmes, 4 novelas, 6 séries, 1 minissérie e 1 caso especial. Essas produções foram pesquisadas na Cinemateca Nacional (www.cinemateca.gov.br) e no Arquivo Memória Globo (memoriaglobo.globo.com).

Segundo Guimarães (2008), partindo da concepção social, materializada e cíclica do conhecimento, o foco recai sobre a busca pela compreensão, organização e representação desse conhecimento para torná-lo disponível a um número maior de pessoas. Nessa perspectiva, pode-se entender que o conteúdo de obras ficcionais narrativas (audiovisuais, literatura e música) pode caracterizar-se enquanto informação que precisa ser organizada para uma melhor recuperação.

3 Resultados e discussões

Toda a obra amadiana e caymminiana estão ligadas à proteção dos “valores populares” através da literatura. Procura manter viva as tradições e práticas do cotidiano e representar o povo como fonte legítima da identidade nacional, de forma geral, e baiana, de forma particular.

Portanto, a concepção de linguagem como representação do mundo é discutida no âmbito da CI como um elemento de organização do conhecimento por meio da teoria do conceito de acordo com o pensamento de Dahlberg (1978) que diz:

> O conhecimento fixou-se através dos elementos da linguagem. Novos conhecimentos aparecem com novos elementos linguísticos e também através destes tornam-se mais claros e distintos. Podemos dizer que este processo de crescimento há de perdurar enquanto o homem existir sobre a terra e utilizar a linguagem como expressão de seus pensamentos.

O uso da linguagem como expressão do pensamento humano se manifesta por meio das relações sociais e os processos de comunicação do sujeito com os seus semelhantes. Essa manifestação pode se desenvolver também nos fenômenos artístico-culturais, que se materializam em artefatos culturais (cinema, dança,

escultura, literatura, música, pintura e teatro). Esses objetos são recursos informacionais de preservação e tutela da memória dos acontecimentos sociais.

O tratamento da informação envolve duas especialidades identificadas como: tratamento descritivo e tratamento temático. O primeiro consiste em identificar os elementos descritivos de um documento, tais como: autor, título, ano de publicação, etc. O segundo recai no propósito de identificar o documento a partir do seu conteúdo temático, ou seja, o assunto.

O processo de identificação do assunto de um documento envolve a extração de conceitos e consequentemente apontam para construção de modelos de representação da informação. Hjorland (1992), ao discutir sobre o conceito de assunto caracterizou suas formas de tratamento e apresentou uma metodologia que se constitui a partir da investigação de domínios específicos do conhecimento por meio da identificação e contextualização de suas condições culturais, históricas e linguísticas.

Tennis (2012) apresenta as definições de escopo e alcance de uma análise de domínio, bem como o propósito da mesma. A respeito do escopo e alcance ele afirma que é necessário especificar a extensão do mesmo, de forma que deve-se nomear o domínio e detalhar suas extensões e exclusões. O propósito da análise está relacionado ao seu gênero (descritiva ou instrumental). Portanto, de acordo com a proposta do autor, a pesquisa apresenta um escopo nas produções audiovisuais oriundas das adaptações das obras literárias de Jorge Amado que apresentam em sua trilha sonora composições de Caymmi com um propósito descritivo e desenvolvido por meio das técnicas de análise de assunto.

Assim, é possível perceber que a materialidade dessas informações se constitui a partir de um fenômeno social, ou seja, a informação passa do estado

intangível (informação como conhecimento) para o estado tangível (informação como coisa), conforme aponta Buckland (1991).

Nessa perspectiva, a pesquisa dá ênfase à técnica de análise de assunto de produções audiovisuais a partir da relação de obras literárias e musicais. Os instrumentos de representação da informação (RI) estão baseados no estudo de Novellino (1996) que aponta a análise de assunto de um documento e seu resultado numa expressão linguística, além da atribuição de conceitos ao documento analisado como os principais passos no processo de RI.

O quadro 1 e 2 apresentam os elementos extratidos da produção audiovisual baseadas na obra de Jorge Amado, que figuram como elementos de representação descritiva dos documentos analisados.

Nº	Filme	Ano	Diretor	Gênero	Livro em que a obra fora baseada	Ano	Duração em min.	Musica(s) de Dorival Caymmi e/ou Jorge Amado que compõe a trilha sonora	
1	Terras violentas	1948	Eddi Bernoudy e Paulo Machado	Drama	Terras do sem-fim	1943	113	Canção	ABC dos marajós
								Compositor/es	Dorival Caymmi
								Intérprete	Dorival Caymmi
2	*The Sandpit Generals*	1971	Hal Bartlett	Drama	Capitães da Areia	1937	102	Canção	Canção da Partida
								Compositor/es	Dorival Caymmi
								Intérprete	Dorival Caymmi
3	Dona Flor e Seus Dois Maridos	1976	Bruno Barreto	Comédia	Dona Flor e Seus Dois Maridos	1966	120	Canção	São Salvador
								Compositor/es	Dorival Caymmi
								Intérprete	Dorival Caymmi
4	Tenda dos Milagres	1977	Nelson Pereira dos Santos	Drama	Tenda dos milagres	1969	132	Canção	Afoxé
								Compositor/es	Dorival Caymmi
								Intérprete	Dorival Caymmi
5	O capeta Carybé	1996	Agnaldo “Siri” Azevedo	Documentário	O capeta Carybé	1986	22	Canção	
								Compositor/es	Dorival Caymmi
								Intérprete	Paulo Moura
6	Quincas Berro D'Água	2010	Sérgio Machado	Comédia/ Chanchada	A Morte e a Morte de Quincas Berro d'Água	1959	102	Canção	Canto De Nanã
								Compositor/es	Dorival Caymmi
								Intérprete	MPB-4

Quadro 1 – Filmes baseados em obras de Jorge Amado

Fonte: Elaborado pelos autores.

Nº	Obra	Natureza do Artefato	Ano	Diretor	Livro de origem do artefato	Ano	Musica(s) de Dorival Caymmi e/ou Jorge Amado que compõe a trilha sonora	
1	Gabriela	Série	1960	Mauricio Sherman	Gabriela, cravo e canela	1958	Canção	Alegre Menina
							Compositor/es	Jorge Amado / Dorival Caymmi
							Intérprete	Djavan
							Canção	Horas
							Compositor/es	Dorival Caymmi
							Intérprete	Quarteto Em Cy
							Canção	Adeus
							Compositor/es	Dorival Caymmi
							Intérprete	Walker
							Canção	Modinha para Gabriela
							Compositor/es	Dorival Caymmi
							Intérprete	Gal Costa
2	Gabriela	Série	1975	Walter Avancini	Gabriela, cravo e canela	1958	Canção	Modinha para Gabriela
							Compositor/es	Dorival Caymmi
							Intérprete	Gal Costa
							Canção	Horas
							Compositor/es	Dorival Caymmi
							Intérprete	Dorival Caymmi
							Canção	Adeus
							Compositor/es	Dorival Caymmi
							Intérprete	Dorival Caymmi
3	Terras do Sem Fim	Novela	1981	Herval Rossano	Terras do Sem-Fim	1943	Canção	Cantiga de Cego
							Compositor/es	Jorge Amado
							Intérprete	Dorival Caymmi
4	Tenda dos Milagres	Série	1985	Guto Graça Mello	Tenda dos milagres	1969	Canção	Afoxé
							Compositor/es	Dorival Caymmi
							Intérprete	Dorival Caymmi
5	Tereza Batista	Série	1992	Paulo Afonso Grisolli	Teresa Batista cansada de guerra	1972	Canção	Vamos Falar de Tereza
							Compositor/es	Danilo / Dorival Caymmi
							Intérprete	Danilo / Dorival Caymmi
							Canção	Modinha para Tereza Batista

							Compositor/es	Dorival Caymmi / Jorge Amado
							Intérprete	Dorival Caymmi
6	Dona Flor e Seus Dois Maridos	Série	1998	Mauro Mendonça Filho	Dona Flor e Seus Dois Maridos	1966	Canção	Caminhos do Mar
							Compositor/es	Dorival Caymmi
							Intérprete	Gal Costa
							Canção	O Bem do Mar
							Compositor/es	Dorival Caymmi
							Intérprete	Dorival Caymmi
							Canção	Saudade de Amar
							Compositor/es	Dori Caymmi / Paulo Cesar Pinheiro
							Intérprete	Nana Caymmi
7	Porto do Milagres	Novela	2001	Marcos Paulo, Roberto Naar	Mar Morto e A Descoberta da América pelos Turcos	1935/1994	Canção	Caminhos do Mar
							Compositores	Dorival Caymmi
							Intérprete	Gal Costa
							Canção	O Bem do Mar
							Compositores	Dorival Caymmi
							Intérprete	Dorival Caymmi
							Canção	É Doce Morrer no mar
							Compositores	Dorival Caymmi
							Intérprete	Dori Caymmi
8	Gabriela	Série	2012	Mauro Mendonça Filho	Gabriela, cravo e canela	1958	Canção	Modinha para Gabriela
							Compositor/es	Dorival Caymmi
							Intérprete	Gal Costa

Quadro 2 – Produções televisivas adaptadas da obra de Jorge Amado

Fonte: Elaborado pelos autores.

4 Conclusão

Entendemos por fim que a interseccionalidade é um conceito teórico-prático aplicável à Ciência da informação, onde no caso analisado o conjunto de informação configura-se como elemento de tutela do registro de um fenômeno social revelando características de um grupo por meio do tratamento temático da informação (Guimarães, 2008). Para tanto, ressalta-se que o objetivo desta pesquisa fora atingido e no âmbito da OC tal estudo serve de subsidio e aponta possibilidades no campo da cultura e das representações sociais.

Referências

ALBUQUERQUE JR, Dúrval Muniz de. **A invenção do Brasil e outras artes.** 4ª ed. São Paulo: Cortez, 2009.

BUCKLAND, Michael K. Information as thing. **Journal of the American Society for Information Science**, v. 45, n. 5, p. 351-360, 1991.

CRENSHAW, K. Documento para o encontro de especialistas em aspectos da discriminação racial relativos ao gênero. **Estudos Feministas**, Florianópolis, v. 10, n. 1, p. 171-188, jan. 2002.

DAHLBERG, Ingetraut. Teoria do Conceito. **Revista Ciência da Informação**, Rio de Janeiro, v.7, n.2, p.101-107, 1978.

FERREIRA, Aurélio Buarque de Holanda. **Miniaurélio Século XXI**: o minidicionário da língua portuguesa. 5.ed. Rio de Janeiro: Nova Fronteira, 2001.

FOUCAULT, Michel. **Microfísica do poder.** Rio de Janeiro: Graal, 2000.

GUIMARÃES, José Augusto Chaves. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for KnowledgeOrganization (ISKO). **Revista Ibero-americana de Ciência da Informação**, v. 1, n. 1, p. 77-99, jan./jun. 2008.

HJORLAND, Birger. The concepto f “subject” in Information Science. **Journal of Documentation**, v.48, n.2, p172-200, June 1992.

MORENO, Antônio do Nascimento. **A personagem homossexual no cinema brasileiro**. 1995. 148 f. Dissertação (Mestrado em Artes) – Instituto de Artes, Universidade Estadual de Campinas, Campinas, 1995.

NOVELLINO, Maria Salet Ferreira. Instrumentos e metodologias de representação da informação. **Informação & Informação**, Londrina, v.1, n.2, p.37-45, jul/dez. 1996.

ROLNIK, Suely. **Cartografia Sentimental:** Transformações contemporâneas do Desejo. São Paulo: Estação Liberdade, 1989.

TENNIS, Joseph T. Com o que uma análise de domínio se parece no tocante a sua forma, função e gênero? **Brazilian Journal of Information Science**, Marília (SP), v. 6, n. 1, p. 3-15, jan./jun. 2012. Disponível em: http://www2.marilia.Unesp.br/revistas/index.php/bjis/index. Acesso em: 30 jul 2013.

A mediação no domínio da organização do conhecimento

Mona Cleide Quirino da Silva Farias
Universidade Estadual Paulista
monaquirino@gmail.com

Carlos Cândido de Almeida
Universidade Estadual Paulista
carlosalmeida@marilia.unesp.br

Daniel Martínez-Ávila
Universidade Estadual Paulista
dmartinezavila@marilia.unesp.br

Introdução

A organização do conhecimento em suas abordagens socioculturais, notadamente em Beghtol (1986, 1995, 2001, 2002, 2005), Campbell (2000, 2010), Dal' Evedove (2014), García Gutiérrez (1998, 2002a, 2002b, 2002c, 2004, 2006, 2008, 2011), Guimarães (2005), Guimarães e Milani (2010, 2011), Guimarães e Pinho (2008), Hjørland (2003, 2007, 2008, 2010, 2013), Hudon (1997, 1999, 2003), Milani (2010), Milani et al. (2009) e Pinho (2006, 2010) entre outros, nos impulsiona a refletir sobre o contexto na representação e organização do conhecimento. Esse contexto, ao fim e ao cabo, constitui-se por processos interpretativos e de mediação.

Vale mencionar que o termo “mediação” é ainda polissêmico não somente na Ciência da Informação. Bortolin (2010) afirma que há uma “pluralidade” de definições

de mediação, por isso não apresenta um conceito consensual, nem talvez científico, que forneça uma base para reflexão de modo amplo e coerente para as diversas áreas e linhas da Ciência da Informação. Para Malheiro e Ribeiro (2011, p.156) "[…] falta à CI o uso conceptual da mediação, entendida como instância articuladora entre diferentes partes sempre em determinadas situações e contextos".

Este artigo objetiva reintroduzir o tema mediação como elemento necessário à compreensão dos processos de representação e organização do conhecimento, mediante a contribuição das dimensões conceitual, cultural e profissional. Através de uma análise da literatura crítica da organização do conhecimento, analisaremos seu estatuto como um campo de mediações em que os processos de representação e organização podem ser reconhecidos como mediação social, cujas atividades promovem, consequentemente, a mediação do conhecimento entre sujeitos tendo em vista seus valores e metas capazes de reconhecer outros pontos de vista e as diferentes comunidades, valores e culturas.

2 A mediação na organização do conhecimento

2.1 A Dimensão Conceitual

A mediação deve ser encarada como um processo social de cunho interpretativo, e, como tal, requer atividades que visem a organização do conhecimento, isto é, que consigam abranger aspectos éticos e socioculturais para o tratamento e a socialização do conhecimento, na medida em que considera os múltiplos contextos e metas das comunidades.

A mediação corresponde a uma ação que trata de conhecer os contextos, culturas, sujeitos e comunidades, no intuito de refletir e repensar os modos de conhecimento, pois o conhecimento é uma constante que se modeliza a todo o momento. Nesse aspecto entendemos que a organização do conhecimento é um campo mediador, onde é necessário conhecer os contextos das comunidades discursivas.

A mediação como um processo interpretativo de ordem social e cultural deve ser analisada a partir do contexto no qual a mesma propõe conhecer e interpretar. É por meio do contexto que podemos conhecer as culturas e as comunidades. Neste sentido, a figura humana é justificada, pois esta traz consigo todo seu referencial de

conhecimento, crenças, ética, linguagens, o que influencia a ação de mediar no processo de representação do conhecimento.

Para tanto, a mediação deve ser também analisada como instrumento de transformação dos sujeitos. O profissional é o interlocutor de tal processo, tendo em vista a sua contribuição evidenciada por meio da organização e representação do conhecimento. A organização do conhecimento é, dessa maneira, considerada um campo cujos seus processos são mediadores do conhecimento. Esse elemento militante das práticas de mediação, tal como está presente em Freire (1976), Almeida Júnior (2003, 2007, 2009) e Bortolin (2010), embora questionável, é fundamental para entender seu compromisso ético e responsabilidade social.

Assim, a mediação é um processo dialético, visto que as relações entre os sujeitos e as inúmeras formas de conhecimento não consistem em algo reificado. No entanto, para representar o conhecimento, qualquer que seja, devemos levar em conta ideias, cosmovisões, contextos, domínios de conhecimentos, experiências, linguagens e a própria noção de cultura.

Entendemos que toda forma de organização do conhecimento é uma mediação entre o conhecimento e as comunidades de usuários. A ideia de usuário faz pouco sentido quando resolvemos assumir o papel social de mediação na representação do conhecimento, pois ao fim e ao cabo, tratamos com comunidades, com sujeitos históricos. Uma classificação é resultado da mediação, assim como o tesauro. Não são simples ferramentas ou produtos da organização do conhecimento, mas emanam da práxis social de mediação, enquanto capacidade propriamente humana. Para tanto, a organização do conhecimento implica um processo contínuo de diálogo.

A mediação é um processo comunicativo e dialógico, portanto, a ação realizada não é uma intervenção neutra a uma realidade externa. O mesmo procede com as comunidades dos sistemas de organização e representação do conhecimento. A acepção funcionalista do “sistema” também cai em desuso com a concepção comprometida de mediação. Na busca por conteúdos informacionais, as comunidades trazem consigo valores, categorias, representações, histórias, linguagens, isto é, cultura no sentido antropológico. Um domínio de conhecimento não suplanta estas variáveis.

A organização do conhecimento, aceitemos ou não, é um campo de mediação. Tomamos a ideia de campo de Bourdieu (1983, 1989, 2004), no sentido que se

entende como um espaço de conflitos, de lutas por posições, de disputas entre formas de representação da realidade social. Os resultados da organização do conhecimento não se operam sobre o vazio. Ao contrário, implicam atividades de análise interpretativa que interferem e, consequentemente, ensejam representar uma comunidade discursiva. A representação do conhecimento é mediação que, segundo Fernández-Molina et al. (2005), pode ser entendida como atividade que busca disponibilizar informação a diversos usuários - leiam comunidades -, sem distinção e com o propósito de que os usuários se sintam refletidos no contexto representado.

Sob tal ideia de mediação, aglutinar-se-iam as tendências pós-modernas e socioculturais da organização do conhecimento, em vários matizes, como a materialização do processo mediativo levado a cabo nos espaços de informação.

2.2 A Dimensão Cultural

A mediação se processa em uma realidade sociocultural, a qual devemos entender com maior cuidado de detalhes. Uma simples enumeração de variáveis mensuráveis não responderia ao conhecimento do contexto. A avaliação de seu comportamento, enquanto a sua função episódica de usuário de um sistema de recuperação da informação, tampouco representa a dimensão social. Uma análise demográfica da comunidade, embora necessária, não esgota o conhecimento do contexto, muito menos o substitui.

Se assumirmos a mediação como instância fundante, o contexto a que se refere é notadamente multicultural: “O processo de representação do conhecimento, assim como os instrumentos que o permeiam e os produtos que dele decorrem não são neutros, estando imbuídos de valores morais” (FERNÁNDEZ-MOLINA et. al., 2005, p. 204).

A mediação nesse sentido é uma ação social, uma intervenção, frente aos contextos multiculturais. O multiculturalismo, conforme Fernández-Molina et al. (2005), traz à sociedade a necessidade de reconhecer e respeitar as diferenças, no intuito de promover a coexistência do respeito entre indivíduos e grupos.

Nesse sentido, García Gutiérrez (2002a) propõe a substituição de profissional da informação por mediador, considerando que esta denominação é mais abrangente, tendo em vista a atuação do mediador em contextos multiculturais.

Para tanto, a mediação multicultural assume uma efetiva complexidade pelo fato de que deve dedicar atenção, tanto ao momento de recepção do conhecimento como também as possibilidades de sua fragmentação, considerando que um dos desafios do multiculturalismo é tentar evitar situações de incomunicabilidade e desse modo buscar novos espaços de mediação e de tradução (FERNÁNDEZ-MOLINA et al., 2005).

Guimarães e Pinho (2008) afirmam que o universo cultural não pode ser reduzido a um paradigma positivista a ponto de não se notar a realidade, deixando prevalecer a visão de posições dominantes. Os autores alertam ainda que os sistemas de representação e organização do conhecimento necessitam possuir, em seus fundamentos, princípios éticos que promovam uma visão transcultural e permitam a divulgação de variados aspectos relacionados ao conhecimento.

O mediador deve buscar compreender os múltiplos contextos e suas linguagens, assim como suas diversas formas de expressão representadas pelos sujeitos de um espaço sociocultural do conhecimento. Nesse aspecto, o profissional deve estar atento para as limitações de ordem linguística na busca de evitar o "imperialismo linguístico", destacado por Hudon (1997).

Nessa discussão, a autora enfatiza sobre os tesauros multilíngues e considera que, para enfrentar as barreiras linguísticas, os tesauros multilíngues podem atuar de modo que possibilitem aos usuários o acesso à informação considerando as culturas e línguas distintas.

Vale ressaltar que a mediação ocorre por meio da manifestação de linguagens utilizadas, por exemplo, em um determinado sistema de representação elaborado por profissionais. Assim, a construção de esquemas de conhecimento deve reforçar a garantia das questões multiculturais que devem ser destacadas pelos profissionais no instante da mediação, na representação. As questões afetas ao multiculturalismo refletem a ética, no que corresponde à defesa de valores e a inclusão no intuito de resguardar os direitos dos cidadãos (FERNÁNDEZ-MOLINA et al., 2005).

A cultura é um complexo de signos que merece atenção quando da elaboração da representação e organização do conhecimento, principalmente por estar ligada à linguagem e, consequentemente, requer um cuidado ético e social. Segundo García Gutiérrez (2004, p. 83), "A cultura, considerada como o imaginário simbólico e

intangível do imaginário, é uma fenomenologia exclusivamente concebida como um sistema aberto, ou seja, pensá-la como um sistema fechado e formal é um erro".

De acordo com esta perspectiva, os profissionais mediadores devem estar predispostos a perceber que no campo da cultura há uma diversidade de códigos, simbologias e linguagens que podem auxiliá-los no processo de diálogo com as comunidades. Tal diálogo estabelece a relação entre profissionais e sujeitos, sujeitos e contextos culturais, profissionais e sistemas de representação, organização do conhecimento e comunidades.

O multiculturalismo não é apenas uma perspectiva, uma abordagem teórica exotérica, alijada da prática, mas é a própria realidade social, a condição de nossas sociedades. Convivemos, assim como argumenta García Canclini (2001), em uma mesma realidade, em um mesmo espaço com múltiplas temporalidades históricas. Se atravessarmos o Brasil, encontramos culturas milenares coexistindo com simulacros europeus de sociedade, no que tange às condições materiais de vida. Essas culturas, que representam a maior parte das mais de 200 línguas faladas no Brasil não são representadas adequadamente por sistemas de organização do conhecimento. Do mesmo modo, se passamos a países europeus, notamos a complexidade que é o planejamento das políticas linguísticas, pois ademais das 24 línguas oficiais da União Europeia, existem muitos outros grupos linguísticos não incluídos. Consequentemente, o multiculturalismo não pode ser usado apenas como discurso pela organização do conhecimento, pois, por detrás dele, há uma realidade cultural híbrida a qual devemos empreender ações de mediação.

Assim, é peremptório considerar a diversidade cultural, tal como o contexto como mencionado por Fernández-Molina et al. (2005) que reconhece a necessidade de explicitação de valores e paradigmas que permeiam os aspectos históricos, sociais e culturais na representação e organização do conhecimento.

Conquanto, há uma questão ética na ação de mediar. Guimarães (2000) categoriza cinco instâncias correspondentes às atribuições éticas que envolvem a atuação profissional: a instância relacionada ao usuário, à organização, à informação, com a profissão e com o próprio profissional. No escopo dessa discussão, tais instâncias reforçam a mediação como um compromisso ético com a comunidade. Assim, abrimos espaço às consequências profissionais da mediação.

2.3 A Dimensão Profissional

A mediação no campo da organização do conhecimento se configura na prática como um processo fundamentalmente interpretativo, que implica ação ética e dialógica sobre o conhecimento como a base está calcada nos contextos, linguagens e na compreensão de cultura. Entendemos que os estudos de cunho sociocultural mencionados anteriormente vislumbram uma prática do profissional, seja realizada pelo bibliotecário, arquivista, museólogo etc., enquanto sujeito mediador da representação e organização do conhecimento.

Sobre a prática do profissional, Guimarães (2005) ressalta que a "dimensão profissional da ética enquanto conjunto de valores que um segmento social, caracterizado pela especificidade de um saber e de um fazer (profissão) estabelece como necessários e fundamentais ao exercício dessa profissão".

Segundo Fernández-Molina e Guimarães (2002), no cerne de discussões tratadas na Ciência da Informação, na temática do profissional vêm sendo tratada com maior destaque as discussões que estão relacionadas à prática profissional, ou ainda uma abordagem sobre os problemas dos direitos específicos, isto é, à especificidade, direito ao *copyright*, direito à liberdade intelectual, censura, entre outros.

Para tanto, destacamos a existência de alguns trabalhos que enfatizam o papel ético do mediador na organização do conhecimento como, por exemplo, no contexto da mídia (GARCÍA GUTIÉRREZ, 2013, 2014; GARCÍA GUTIÉRREZ; MARTÍNEZ-ÁVILA, 2014a) e também ao que corresponde à formação crítica do profissional em organização do conhecimento (GARCÍA GUTIÉRREZ; MARTÍNEZ-ÁVILA, 2014b).

Segundo Fernández-Molina et al. (2005) o profissional da informação tem a responsabilidade de assegurar a diversidade de acesso às informações culturais. Nesse sentido, a sua atuação concerne na projeção, avaliação, manutenção e revisão dos sistemas de representação. Assim, a mediação é um processo sociocultural de caráter híbrido relacionado à prática dos profissionais na organização do conhecimento.

O profissional é um ser social que reflete, dialoga e interage com os múltiplos contextos das diversas comunidades. No âmbito dos processos, assim como das atividades de organização do conhecimento, Guimarães e Pinho (2008) consideram que a representação implica em uma atividade intelectual e que exige do profissional

uma postura consciente e crítica sem deixar escapar um conhecimento sobre aspectos históricos e sociais que circundam o conhecimento registrado e socializado.

Compreendemos que no âmbito da organização do conhecimento já se evidenciam interesses com as questões profissional e de formação de pessoal para atuação como mediador. Nessa linha, devemos destacar os trabalhos que reconfiguraram a visão das técnicas, dos produtos e instrumentos de organização do conhecimento, os quais agora, sob uma visão crítica e de não neutralidade, motivam os profissionais a trabalharem por uma sociedade que reconheça o fato multicultural e o papel semiótico que representam na organização do conhecimento.

A dimensão profissional da mediação na organização do conhecimento tem como base o compromisso ético do profissional, desde a representação de um conteúdo, construção de um sistema de classificação até o reconhecimento da necessidade de aprender métodos sociológicos, antropológicos e linguísticos para proporcionar uma mediação mais integrada.

Conclusão

Em resumidas contas, amparamo-nos nas palavras de García-Gutiérrez e Martínez-Ávila (2014a), pois assumem o reconhecimento teórico da presença subjetiva e dos vieses da organização do conhecimento como mediação, um trabalho que geralmente era considerado neutro, banal e mesmo objetivo e que pode ser aplicado na prática através de uma formação crítica dos profissionais. Como guisa de conclusão, devemos listar resumidamente os argumentos manifestos até aqui.

Em primeiro lugar, no âmbito dos sistemas de representação é importante o reconhecimento da não neutralidade por parte dos instrumentos e dos profissionais que os gerenciam, tal como já apontam as perspectivas de Hjørland (2003), Mai (2004) e Olson (2002). Desse modo, é necessário dizer que nos instrumentos de representação como, por exemplo, tesauros, sistemas de classificação facetada, linguagens documentárias, estão contidos valores e em alguns casos posicionamento dos profissionais envolvidos.

Em segundo lugar, a organização do conhecimento deve ser um espaço crítico e interpretativo que favoreça a prática da mediação, tendo em vista uma atuação ética, multicultural e adequada às características das diversas comunidades que fazem uso desse conhecimento.

Um terceiro argumento é que se deve, por consequência lógica, substituir, tal como sugere García Gutiérrez (2002a), a ideia de profissional da informação ou especialista em sistemas de organização do conhecimento - que nos parece mais a descrição de uma função ou tarefa -, pela de mediador, considerando a abrangência da atuação em contextos multiculturais.

Por fim, a organização do conhecimento não é, em absoluto, um tipo de mediação, ou um caso específico entre as tarefas do profissional da informação, é mais bem equivalente à mediação. Nesse sentido, já seria justo considerar que trabalhamos em mediação do conhecimento e não apenas na organização instrumental deste.

Referências

ALMEIDA JÚNIOR, O. F. Mediação da informação: discutindo a atuação do bibliotecário. In: FADEL, Barbara (Org.). **A informação nas organizações sociais**: desafios em face de multiplicidade de enfoques. Marília: FUNDEPE, 2003. 1 CD-ROM.

_____. Mediación e información. **Ibersid,** p. 27-35, 2007.

_____. Mediação da informação e múltiplas linguagens. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 2, n. 1, p. 89-103, jan./dez. 2009. Disponível em: <http://inseer.ibict.br/ancib/index.php/tpbci/article/view/17/39>. Acesso em: 21 de julho de 2015.

BEGHTOL, C. Semantic Validity: concepts of warrant in bibliographic classification systems. **Library Resources & Technical Services.** v. 30, n. 2, p. 109-125, 1986.

BEGHTOL, C. Domain analysis, literary warrant, and consensus: the case of fiction studies, **Journal of the American Society for Information Science and Technology**, v. 46, n. 1, p. 30-44, 1995.

BEGHTOL, C. Relationships in Classificatory structure and meaning. In: C.A. Bean & R. Green, (Eds.). **Relationships in the organization knowledge.** Dordrecht, Netherlands: Kluwer, 2001, p. 99-113.

BEGHTOL, C. Universal concepts, cultural warrant, and cultural hospitality. In: LÓPEZ-HUERTAS, M. J. (Ed.). **Challenges in knowledge representation and organization for the 21 centurry:** integration of knowledge across boundaries. Wurzburg: Ergon-Verlag. 2002. p. 45-49. (Advances in Knowledge Organization, 8).

BEGHTOL, C. Ethical decision-making for knowledge representation and organization systems for global use. **Journal of the American Society for Information Science and Technology**, v. 56, n. 9, p. 903-912, 2005.

BORTOLIN, Sueli. **Mediação oral da literatura**: a voz do bibliotecário lendo ou narrando. 2010. 232f. Tese (Doutorado em Ciência da informação)- Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

BOURDIEU, P. O campo científico. In: ORTIZ, R. (Org.). **Pierre Bourdieu**: sociologia. São Paulo: Ática, 1983.

BOURDIEU, P. **O poder simbólico**. Lisboa: DIFEL; Rio de Janeiro: Bertrand Brasil, 1989.

BOURDIEU, P. **Os usos sociais da ciência**: por uma sociologia clínica do campo científico. São Paulo: Unesp, 2004.

CAMPBELL, G. Queer theory and the creation of contextual subject access tools for gay and lesbian comunities. **Knowledge Organization,** v.27, n. 3, p. 122-131. 2000.

CAMPBELL, G. Tensions between language and discourse in North American knowledge organization, **Knowledge Organization**, v. 37, n. 1, p. 51-57. 2010.

DAL' EVEDOVE, P. R. **O tratamento temático da informação em abordagem sociocultural:** diretrizes para definição de política de indexação em bibliotecas universitárias. 2014. 259f. Tese (Doutorado em Ciência da Informação) Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

FERNÁNDEZ-MOLINA, J. C.; GUIMARÃES, J. A. C. Ethical aspects of knowledge organization and representation in the digital environment: their articulation in professional codes of ethics. In: LÓPEZ-HUERTAS, M. J. (Ed.). **Challenges in knowledge representation and organization for the 21 st century**: integration of knowledge across boundaries.Würzburg: Ergon-Verlag, 2002. p. 487-492. (Advances in Knowledge Organization, 8).

FERNÁNDEZ-MOLINA, J. C.; GUIMARÃES, J. A. C.; VIDOTTI, S. A. B. G.; FLAMINO, A. N.; SOUZA, A. S.; CAMARGO, L. S. A.; SILVA, M. S.; MORENO, P. S.; RAMALHO; R. A. S. Aspectos éticos de las nuevas tecnologías de la información y la comunicación y su reflejo en la organización del conocimiento. In: GASCÓN, J.; BURGUILLOS, F.; PONS, A. (Org.). **La dimensión humana de la organización del conocimiento**. Barcelona: Universitat de Barcelona, 2005. p. 177-186.

FREIRE, P. **Ação cultural para a liberdade e outros escritos**. São Paulo: Paz e Terra, 1976.

GARCÍA CANCLINI, N. **Culturas híbridas**: estrategias para entrar y salir de la modernidad. Buenos Aires: Paidos, 2001.

GARCÍA GUTIÉRREZ, A. **Principios de lenguaje epistemográfico**: la representación del conocimiento sobre Patrimonio Histórico Andaluz. Sevilla: Instituto Andaluz del Patrimonio Histórico, 1998. (Cuadernos técnicos, 3).

_____. Knowledge organization from a “culture of the border”: towards a transcultural ethics of mediation. In: LÓPEZ-HUERTAS, M.J. (Ed.).**Challenges in knowledge representation and organization for the 21st century:** integration of knowledge across boundaries. Würzburg: Ergon-Verlag, 2002a., p. 516-522. (Advances in Knowledge Organization, 8).

_____. **La memoria subrogada**: mediación, cultura y conciencia en la red digital. Granada: Editorial de la Universidad de Granada, 2000b.

_____. Exomemoria y cultura de frontera: hacia una ética transcultural de la mediación. In: **VII Congreso Internacional sobre Organización del Conocimiento celebrado en Granada** (España), 2002c.

_____. **Otra memoria es posible**: estrategias descolonizadoras del archivo mundial. Sevilla: Universidad de Sevilla, 2004.

_____. Cientificamente favelados: uma visão crítica do conhecimento a partir da epistemografia. **Transinformação,** Campinas, v.18, n.2, p.103-112, maio/ago. 2006.

_____. **Outra memória é possível:** estratégias descolonizadoras do arquivo mundial. Petrópolis: Vozes, 2008.

_____. Desclassification in knowledge organization: a post-epistemological essay. **TransInformação**, Campinas, v.23, n. 1, p. 5-14, jan./abr., 2011.

_____. Entrevistas periodísticas y textos declarativos: un procedimiento de análisis y recuperación documentales basado en modos de búsqueda pronominal. **El profesional de la información**, v. 22, n. 4, p. 315-325, 2013.

_____. Análisis documental de noticias de prensa en sistemas de información factual. **Revista española de documentación científica**, v. 37, n. 2, e046, 2014.

GARCÍA-GUTIÉRREZ, A.; MARTÍNEZ-ÁVILA, D. Critical Organization of Knowledge in Mass Media Information Systems. **Knowledge Organization**, v. 41, n. 3, p. 205-216, 2014a.

GARCÍA-GUTIÉRREZ, A.; MARTÍNEZ-ÁVILA, D. Critical Organization of Knowledge in Mass Media Information Systems. **El profesional de la información**, v. 23, n. 5, p. 493-500, 2014b.

GUIMARÃES, J. A. C. O profissional da informação sob o prisma de sua formação. In: VALENTIM, M. L. P. (Ed.). **Profissionais da informação**: formação e atuação profissional. São Paulo: Polis, 2000. p. 53-70.

_____. Aspectos éticos em organização e representação do conhecimento (O.R.C): uma reflexão preliminar.Conferência apresentada na mesa- redonda: Organização do conhecimento e interdisciplinaridade. In: **Seminário: Memória, informação e organização do conhecimento**: cruzando fronteiras da identidade. Rio de Janeiro, Unirio, 25 e 26 de agosto de 2005. Marília, 2005. 24p.

GUIMARÃES, J. A. C.; MILANI, S. O. Bias in the indexing languages: theoretical approaches about feminine issues. En: GNOLI, C.; MAZZOCCHI, F. (Eds.). **Paradigms and conceptual systems in knowledge organization**. Italy: Ergon Verlag. 2010. p. 424- 429. (Advances in Knowledge Organization, 12).

_____. O. Problemas éticos em representação do conhecimento: uma abordagem teórica. **DataGramaZero**, v.12, n.1, fev./2011.

GUIMARÃES, J. A. C.; PINHO, F. A. Reflexiones acerca de las teorías sobre ética en la representación del conocimiento. **Biblios**. Revista de bibliotecnoloría y Ciencias de la Información . Jul./set., p. 1-12, 2008.

HJØRLAND, B. Fundaments of knowledge organization. **Knowledge Organization**, v.30, n. 2, p.87-111, 2003.

_____. Semantics and Knowledge organization. **Annual Review of Information Science and Technology**, v.41, n.1, p. 367-405. 2007.

_____. What is Knowledge organization (KO)?. **Knowledge Organization**, v. 35, n.160 2/3 p. 86-101. 2008.

_____. Concepts, paradigms and knowledge organization. In: GNOLI, C.; MAZZOCCHI, F. (Eds.). **Paradigms and conceptual systems in knowledge organization.** Italy: Ergon Verlag. 2010. p. 38-42. (Advances in Knowledge Organization, 12).

HJØRLAND, B. Theories of knowledge organization-Theories of knowledge. **Knowledge Organization**, v. 40, n. 3, p 169-181. 2013.

HUDON, M. Multilingual thesaurus construction: integrating the views of different cultures in one gateway to knowledge and concepts. **Knowledge Organization**. v. 24, n.2, p.84-91. 1997.

_____. Accessing documents and information in a world without frontiers. **The Indexer,** London, v. 21, n. 4, p. 156-159, 1999.

_____. True and tested products: thesauri on the web. **The Indexer,** London, v. 23, n. 3, p. 115-119, 2003.

MAI, J.-E. Classification in context: relativity, reality, and representation. **Knowledge Organization.** v.31, n. 1, p. 39-48. 2004.

MALHEIRO, A.; RIBEIRO, F. **Paradigmas, serviços e mediação em Ciência da Informação**. Recife: Nectar, 2011.

MILANI, S. O. **Estudos éticos em representação do conhecimento**: uma análise da questão feminina em linguagens documentais brasileiras. 2010. 141 f. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

MILANI, S.; GUIMARÃES, J. A. C.; PINHO, F. A; MOLINA, C. F. Os desvios na representação do conhecimento em um contexto multicultural: abordagens teóricas. En: LLORET ROMERO, N. (Ed.) **Nuevas perspectivas para la difusión y organización del conocimiento**: actas del IX Congreso ISKO España. Valencia: Universidad Politécnica de Valencia, Servicio de Publicaciones, 2009. p. 204-218.

OLSON, H. A. **The power to name**: locating the limits of subject representation in libraries. Dordrecht: Kluwer Academic Publisher, 2002.

PINHO, F. A. **Aspectos éticos em representação do conhecimento**: em busca do diálogo entre Antonio García Gutiérrez, Michèle Hudon e Clare Beghtol. 2006. 123 f. Dissertação

(Mestrado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2006.

_____. **Aspectos éticos em representação do conhecimento em temáticas relativas à homossexualidade masculina**: uma análise da precisão em linguagens de indexação brasileiras. 2010. 149 f. Tese (Doutorado em Ciência da Informação) - Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

O valor da informação e da linguagem na sociedade de consumo

Luciana de Souza Gracioso
Universidade Federal de São Carlos
luciana@ufscar.br

1 Introdução

Sem muitas restrições geográficas e culturais, nos situamos em uma sociedade de consumo e, atualmente, entre a vontade de consumo e do consumo propriamente, há a intermediação das tecnologias de informação. As vontades de consumo de produtos, serviços de informação, são estimuladas nas redes virtuais, assim como as vontades de espelhamento, de construção de identidades e os desejos de pertencimento a grupos. Neste ambiente, temos nos perguntado sobre os posicionamentos da Ciência da informação e das ferramentas de organização do conhecimento, vislumbrando cogitar a sinalização de orientações que promovam processos de recuperação de informação significativos. Assim este trabalho se vincula ao tema que trata da dimensão social e política da organização do conhecimento.

Neste momento inicial é necessário dar destaque a relevância dos trabalhos de B. Hjorland, J. T. Tennis, M. J. López Huertas, M. Hudon, W. Mustafá El Hadi, J. D. Campbell, H.Olson, C. Begthol, A.Garcia Gutierrez, dentre outros, que se dedicam as questões sociais da organização do conhecimento. Este estatuto científico conferido a tais autores foi mapeado e apresentado por Guimarães, em recente conferência ministrada no Colóquio: Organização do conhecimento no tempo: passado, presente e futuro, organizador pelo PPGCI (IBICT/UFRJ) em 2015. No panorama latino americano temos notadamente alguns autores que se ocupam dos estudos sobre os usos sociais da linguagem considerando suas implicações para a organização do conhecimento, dentre eles: S. P. Mostafa, M. A. R. Rojas, M. N. G. Gomez, G. S. Saldanha, M. S. Novelino, E. G. Del Orrico. (SALDANHA, GRACIOSO, 2014).

Sobre os estudos que envolvem a discussão sobre o valor de consumo da informação temos pesquisas que transitam entre o Mercado da informação, anunciadas por Aldo Barreto há 15 anos até aS mais recentes, desenvolvidas por Moura e Gomes (2014) sobre o comércio e o consumo em tempos de conexões digitais.

Já, para o percurso argumentativo que nos propusemos para este trabalho, seguimos um caminho teórico pautado na sociologia e na filosofia da linguagem pragmática, que pretendemos poder contribuir para as discussões sobre a organização do conhecimento em sua esfera social e política no contexto brevemente anunciado.

2 Consumo ou uso da informação?

Os mecanismos de busca atuais advogam e orientam uma prática bilateral de busca e recuperação de conteúdos. Graças aos recursos sintáticos e semânticos da Web, recuperamos. Porém já somos reféns do efeito *filtro bolha* (conceito atribuído a personalização extrema de conteúdos on line de acordo com o perfil de busca dos usuários). A quarta lei da biblioteconomia apresentada por S. Ranganathan, *Poupe o tempo do leitor,* (1931) parece ter sido desvirtuada pelos desenvolvedores iniciantes dos SEO (*Search Engine Optimization*), embora mais recentemente eles tenham se dedicado a produção de uma web com mais conteúdo. As ações de busca da informação na Web são monitoradas a cada segundo e o uso de palavras e expressões de busca nunca foi tão observada pelos gestores de websites. Nosso perfil de usuário da informação tem se convertido para um perfil de consumidores da informação.

Este fluxo de consumo tem desvirtuado a própria organização social e a construção de identidades. Z. Bauman, mais pontualmente seu livro *Modernidade líquida* de 2001, tem nos ajudado a refletir sobre isto. A leitura de Z. Bauman tem sido feita e apropriada em diferentes perspectivas na Ciência da informação. Bernett e Vavakis (2001) discutem “(...) os desafios das tecnologias de informação e comunicação sob a perspectiva da gestão do conhecimento na sociedade em redes (...)”. Barreto, em 2008, resenha a obra: *Comunidade - A Busca por segurança no mundo atual,* escrita por Bauman, que analisa os modelos atuais de construção de

comunidades. Pautados na obra *Modernidade líquida*, Silva e Lopes discutem a *desintermediação da informação* e apresentam as problematizações que surgem para a Biblioteconomia sobre o uso de métodos de regulação e representação de conteúdos. Com abordagem semelhante, Pedro Demo também analisa as *ambivalências da sociedade da informação*, em 2000. Ziller e Moura, em 2010, discutem a *Semiose e fluxos informacionais*: os agenciamentos coletivos e a condição de usuário em ambientes digitais e Mantovani e Moura, ponderam sobre a *informação, interação e mobilidade* em 2012. Albagli, em 2011 analisa *a questão do outro: do território ao digital* e apresenta um interessante discussão sobre os interesses do estado-nação em criar a ilusão sobre a liberdade para criação de identidades. Esta autora também identifica em Habermas, na sua obra *Inclusão do outro* (2002), a perspectiva do pensador em priorizar o lugar do indivíduo cidadão, consciente e conhecedor de seus direitos e deveres, na sociedade. Albagli, ao se debruçar sobre a obra de Habermas (2002) identifica que o autor teme que com a fragilização Estado-nação, *(...) vigoraria uma "lógica de redes" interconectadas anonimamente e supostamente regulada pela mão invisível da economia global, em que as empresas multinacionais são o modelo de conduta, e em que se teria a ausência de "um universo de significados compartilhados intersubjetivamente".* (HABERMAS, 2002, p. 144).

Maia (2001) é outra autora que se volta para este espectro e sinaliza que, antes de se pensar a operacionalização de modelos ou sistemas que potencializem a comunicação virtualizada, a que se pensar em como incentivá-las coerentemente no plano real. "Para fortalecer a democracia, são necessárias não *apenas* estruturas comunicacionais eficientes, ou instituições propícias à participação, mas também devem estar presentes a motivação correta, o interesse e a disponibilidade dos próprios cidadãos para se engajar em debates." (MAIA, 2001). Com este argumento de Maia, seguimos para concatenar nossa proposta investigativa de problematizar sobre como os sistemas de informação e estratégias de organização do conhecimento na Web poderiam ser configurados para intervirem nos processos de busca e recuperação da informação de modo a promover maior significado a este processo. Para sustentar a nossa proposição final, atravessaremos brevemente a rota da Filosofia da linguagem pragmática, com destaque a Teoria dos Atos de Fala de J. Austin.

3 Filosofia da linguagem pragmática: do contexto de uso aos atos de fala nos sistemas de informação

A Filosofia da linguagem contempla um conjunto amplo e aberto de estudos. Demos ênfase nesta pesquisa para a filosofia da linguagem pragmática, ou pragmatismo. Esta corrente filosófica teve suas origens tanto nos Estados Unidos (F.C. S. Schiller, W. James, C. Peirce, J. Dewey, G. H. Mead) como na Inglaterra, mais pontualmente vinculados a Escola de Oxford (J. Austin, G. Ryle, P. Strawson, P. Grice. J. Searle). Os pensamentos de L. Wittgenstein, em seu segundo momento filosófico, teriam impulsionado este movimento europeu. Nas distinções de Marcondes (2006) há duas linhas para desenvolvimento da pragmática inglesa, sob as quais nos apoiamos neste estudo: a concepção contextualista (L. Wittgenstein, 1889-1951) e a concepção performativa (J. Austin, 1911-1960).

Para Austin, dizer é fazer. “A determinação do significado só pode ser feita a partir da consideração do ato que está sendo realizado quando essas expressões são proferidas e das regras que tornam possível a realização desses atos” (MARCONDES, 2006, p. 220). Trata-se da concepção performativa da linguagem em que a “linguagem é basicamente uma forma de realização.” Os atos de fala constituem, segundo os etnógrafos da comunicação (John J. Gumperz e Delleds Hymes), a unidade básica de comunicação verbal, logo, um elemento central ao comportamento humano o que direcionará a sua forma de atuar e agir no mundo.

Sobre isto, Austin elaborou o que ele chamaria de Teoria dos Atos da Fala com o intuito de sistematizar os fenômenos pragmáticos relacionados à linguagem. Para o autor, os atos de fala visariam produzir o sucesso na comunicação. Mais uma vez, seria reforçado o entendimento de que a linguagem não é só usada de maneira constatativa (para descrever fatos) e sim de maneira performativa (para realizar algo). Segundo sua teoria, o ato de fala seria a unidade básica de significação e poderia ser diferenciado em atos locucionários (palavras e sentenças de uma língua), ilocucionários (força do ato da fala) e perlocucionário (consequência dos atos de fala). Por conta disso, um ato de fala seria consequência da relação entre a intenção comunicativa do falante (psicológicas e subjetivas) e as convenções sociais em que ela estará inserida (com diferentes graus de formalidades). Com relação aos diferentes graus de convenções sociais, o autor considera que as regras de

constituição e funcionamento são implícitas, normativas, estabelecidas por condutas sociais previamente estabelecidas e que, diante disso, não seria possível delimitar e analisar todas as regras envolvidas nessas convenções. Nesse sentido, Austin propõe que as análises sobre as convenções sociais deveriam focalizar as falhas ocorridas no uso das regras já implícitas. O cerne de sua teoria estaria nas forças ilocucionárias que impulsionam uma ação de comunicação e as define como forças verídicas, exercitivas, compromissivas, comportamentais e expositivas. (GRACIOSO, 2008).

A partir disto entendemos que os atos de fala movem o real, constroem ações. O mercado de consumo já se atentou para isto e se utiliza, massivamente de recursos da linguagem, mediadas por tecnologias virtuais, para promoverem ações de compra e de consumo. Pensamos então poder vislumbrar alguma intervenção da Ciência da informação, na pragmática informacional da Web, a partir da orientação ao uso da linguagem nas ações de comunicação, uma vez que estas ações se materializam em práticas da vida. Isto reforçaria o posicionamento de Habermas que defende a função propedêutica da teoria da ação comunicativa enquanto teoria de ação social, que por sua vez enfatiza o papel do uso da linguagem como constitutiva da ação comunicativa e esta como constitutiva das ações sociais.

4 Resultados e Considerações finais

A convergência entre ações de uso da linguagem e significação relacionadas as ações práticas da vida cotidiana podem nos ajudar a compreender o sucesso sobre o uso da Web nas investidas do capitalismo para incentivo ao consumo. A Internet, de modo geral, apresenta-se como um território neutro, que pode servir de instrumento ao ataque e a defesa, ao consenso e ao desacordo, ao esquecimento e a ação no mundo da vida. Reconhecendo esta potencialidade da Web e munidos de estratégias que se utilizam de atos ilocucionários imperativos para orientar vontades e individualidades, o mercado tem sim conseguido, através do uso da linguagem na rede, direcionar ações de consumo nas rotinas cotidianas mais elementares. Considerando o cenário exposto por Bauman, sobre a diluição moderna do modelo de Estado, do modelo das Instituições e até mesmo o modelo das relações afetivas e, reconhecendo a imersão gradativa da sociedade no uso da Internet como principal canal de expressão, comunicação e estabelecimento de relações que transitam desde

as afetivas até as de consumo, é que procuramos entender e intervir, neste território, no sentido de procurar refletir sobre o uso de alguns recursos de informação que permitam a esta sociedade, agregar valor e sentido as suas ações de comunicação, tanto as que se utilizam de tecnologias ou não, pois todas têm em maior ou menor grau, relação direta com as práticas cotidianas da vida.

Em um plano mais aplicado, entendemos que profissionais da informação poderiam assumir como parte de suas práticas junto aos desenvolvedores de sistemas de busca e recuperação da informação, além de orientarem sobre toda sistematização, organização e representação de conteúdos já normalizada em outros momentos, orientar também a modelagem de softwares interativos que promovessem e estimulassem, antes mesmo da interação entre usuários e agentes de informação, o questionamento sobre os conteúdos buscados.

As lógicas computacionais em vigor, já são capazes de mapear e até antever “o que” os usuários têm procurado em sistemas de informação, mas nós temos que prover, em nossos sistemas, um espaço para que estes mesmos usuários questionem o “Por quê?” de suas buscas e interação na rede, estimulando-os a responderem a esta mesma pergunta, sobre os conteúdos que receberão como retorno em suas ações de pesquisa e interação na Web.

Concordando com as teorias de J. Habermas, e L. Wittgenstein, consideramos que a validação das ações de informação (comunicação, interação) que ocorrem em nosso cotidiano, precisam se estender do plano dos argumentos a ponto de serem aplicáveis ao campo da vida. Deste modo, se pudermos compreender melhor os desdobramentos dos atos de fala, tal como discriminados por Austin, para orientar a agregação de valor aos atos ilocucionários que proferimos na Web (aqueles que ocorrem nas trocas de informação digitais em processos de comunicação mediados pela tecnologia e que imperativamente nos mandam fazer coisas), poderemos também intervir nas ações práticas da vida, estimulando o consumo de bens culturais e do conhecimento por exemplo. Os usos dos conteúdos buscados e recuperados, se serão empíricos ou abstratos, não vem ao caso. A construção deste valor informativo sobre suas buscas se dará no exato momento em que o próprio buscador utilizar o conteúdo que irá localizar.

Nota: Parte das considerações finais deste artigo assim como detalhes complementares a pesquisa ora apresentada poderão ser conferidos em GONZÁLEZ

DE GÓMEZ, Maria Nélida; RABELLO, Rodrigo (Org.). *Informação*: agentes e intermediação. Prefácio de Rafael Capurro. Brasília: IBICT, 2015. No prelo.

Referências

ALBAGLI, S. A questão do outro: do território ao Digital. In: GOMEZ, M. N. G. de; LIMA, C. R. M. [orgs.]. *Informação e democracia:* a reflexão contemporânea da ética e da política. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2010.

AUSTIN, J.L *How to do Things witch words.* Oxford. Oxford University Press, 302p.

BARRETO, A. A. Recensão: Comunidade: a busca por segurança no mundo atual. DataGramaZero, *v.9 n.1 dez/08.*

BARRETO, A. O mercado da informação no Brasil. *Revista Inf.& Inf.,* v.25, n.1, 2000.

BAUMAN, Z. *Modernidade líquida*. Rio de Janeiro: Zahar Ed., 2001.

BERNETT, D. VARVAKIS, G. Desafios das tecnologias de informação e comunicação sob a perspectiva da gestão do conhecimento na sociedade em redes. DataGramaZero - Revista de Ciência da Informação - v.11 n.3 jun/10.

DEMO, P. Ambivalências da sociedade da informação. *Ciência da informação,* Brasília, v. 29, n. 2, 2000.

GRACIOSO, L. de S. *Filosofia da linguagem e Ciência da Informação*: jogos de linguagem e ação comunicativa no contexto das ações de informação em tecnologias virtuais. 2008. Tese (Doutorado em Ciência da Informação) - Instituto Brasileiro de Informação em Ciência e Tecnologia; Universidade Federal Fluminense, Rio de Janeiro. 2008.

LOPES, M. I. SILVA, E. L. A internet, a mediação e a desintermediação da informação. *DataGramaZero*, v. 12, n. 2, 2011.

MAIA, R. M. *Democracia e a internet como esfera pública virtual:* aproximando as condições do discurso e da deliberação. Disponível em: http://vsites.unb.br/fac/comunicacaoepolitica/Rousiley2001.pdf.

MARCONDES, D. A teoria dos atos de fala como concepção pragmática da linguagem. *Filosofia UNISINOS*, v.7, n. 3, p. 217-230, 2006.

MOURA; M. A.; GOMES, C. A. Comércio e consumo em tempo de conexões digitais: dimensões informacionais. *DataGramaZero* - Revista de Informação - v.15 n.2, 2014.

SALDANHA, G. S.; GRACIOSO, L. S. Filosofia da linguagem e Ciência da Informação na América Latina: apontamentos sobre pragmática e linguagem ordinária. In: RENDON, M. A. (coord.) *El problema Del linguage em la bibliotecologia, ciência de la informacion, documentacion*. Un acercamiento filosófico-teórico. México: UNAM. Instituto de investigaciones bibliotecologicas e de La informacion, 1994.

ZILLER, J.; MOURA, M. A. Semiose e fluxos informacionais: os agenciamentos coletivos e a condição de usuário em ambientes digitais. *Liinc em Revista*, v.6, n.2, setembro, 2010.

Formação acadêmica do docente em descrição arquivística

Laura Maria do Rego
Universidade Estadual Paulista
lauramaria@marilia.unesp.br

José Augusto Chaves Guimarães
Universidade Estadual Paulista
guima@marilia.unesp.br

Natália Bolfarini Tognoli
Universidade Estadual Paulista
nataliatognoli@marilia.unesp.br

1 Introdução

As atividades de organização e representação dos documentos de arquivo têm seu foco no fundo de arquivo, ou seja, no conjunto de documentos produzidos e/ou recebidos por uma mesma pessoa física ou jurídica. O conhecimento gerado a partir do estudo desse fundo é o que denominamos conhecimento arquivístico.

Uma das principais atividades de representação desse conhecimento é a descrição arquivística, cuja função é informar acerca do conteúdo e do contexto dos documentos e de seus elementos formais, com o intuito de dar acesso às informações contidas nos fundos, grupos, séries ou peça documental.

A descrição é, portanto, o ponto mais alto da representação indo ao encontro da finalidade da documentação, que é informar. A tarefa principal do fazer arquivístico

é uma consequência da organização dos fundos para disponibilizar os documentos da melhor forma possível a serviço da sociedade. Sendo assim, o trabalho descritivo tem por objetivo “tornar acessíveis os fundos documentais”. (CRUZ MUNDET, 2001, p.255).

Para Rodrigues (2003, p. 217), a descrição é uma “[...] função-chave no que diz respeito à representação das informações arquivísticas e na possibilidade de acesso a elas”, aliando-se a observação de Tognoli (2012, p.82), no sentido de que “[...] a representação vai além do documento, perpassando suas funções e seus órgãos produtores”.

A descrição, inserida no processo de gestão documental, tem por objeto a elaboração de representações de documentos ou de grupos documentais arquivísticos, onde sua função se estabelece na identificação e representação de atributos que possam “facilitar o acesso, pelos usuários, à informação contextualizada de seus conteúdos”, relata (HEREDIA HERRERA, 2011, p.8).

Nesse sentido, Bellotto (2004, p.177) destaca a possibilidade de se avaliar um profissional com base no trabalho que ele desenvolve em suas atividades de organização documental, já que para a autora, uma elaboração bem feita dos instrumentos de pesquisa, refletirá em uma recuperação satisfatória das informações pelo usuário/pesquisador, tendo em vista que um acervo mal organizado poderá arruinar o trabalho do mesmo. Sendo assim, o arquivista, como profissional qualificado, deverá atuar como gestor dos arquivos, desenvolvendo em sua prática profissional, as atividades de descrição, considerada nuclear na representação.

No entanto, para que essa prática profissional seja exercida de forma sistemática, consistente e metodologicamente defensável, é necessário que a formação educacional desse profissional esteja atenta para garantir elementos suficientes para que se possam tratar os documentos de arquivo de maneira que os mesmos estejam acessíveis aos interessados em suas informações, respeitando sempre todos os princípios da arquivística e fornecendo a esses profissionais conhecimentos necessários sobre sua conduta ética como profissional.

As preocupações no Brasil, quanto à criação de cursos que tivessem como objetivo capacitar profissionais para que os mesmos se tornassem aptos ao tratamento de documentos arquivísticos, começaram a surgir em 1911, com a aprovação da regulamentação do Arquivo Nacional (AN), instituído por meio do Decreto n. 9.197, de 09 de dezembro de 1911. (MARQUES, 2008).

Os arquivistas passaram a se profissionalizavam por meio de cursos oferecidos em programas de Biblioteconomia, e pelo Curso Permanente de Arquivos (CPA), oferecido no AN, com intuito de atender as necessidades da própria instituição, porem essa realidade começa a mudar com a implantação dos primeiros cursos acadêmicos em Arquivologia em 1977. (MARQUES, 2007).

Desde então diversos cursos foram implantados no país, nas mais diversas regiões, com a responsabilidade de proporcionar ao profissional arquivista, uma formação multidisciplinar, uma vez que temos profissionais graduados em Arquivologia ou vindo de outras áreas afins, como a Biblioteconomia ou a História.

Com a expansão dos cursos de graduação em Arquivologia, notadamente nos últimos vinte anos, as Universidades passaram a requerer profissionais com formação específica e experiência em Arquivologia, ora com graduação, ora com pós-graduação na área. Porem a estrutura dos cursos de Arquivologia, nem sempre leva em consideração a especificidade da formação acadêmica para realizar a atribuição de disciplinas entre os docentes, o que leva a situações de duplicidade ou de pulverização da produção científica que se situa entre a sua formação em uma especificidade temática, e que muitas vezes não estabelece um diálogo entre ambas.

Sendo assim, com o intuito de discutir a articulação entre a formação acadêmica e a produção cientifica dos docentes do curso de Arquivologia no Brasil, especificamente daqueles relacionados à temática de descrição, a partir da triangulação desenvolvida por Guimarães (2003), prática de ensino, formação acadêmica e produção cientifica - porém na presente pesquisa abordaremos apenas a formação acadêmica e a produção cientifica - o presente estudo busca verificar a formação e a produção científica dos docentes dos cursos de Arquivologia, especificamente nas disciplinas que tratam a temática descrição, a fim de verificar até que ponto a especificidade temática da formação do docente está relacionada com a sua formação pedagógica e sua produção científica.

2 Docentes de descrição arquivística nos cursos de arquivologia no Brasil: um estudo a partir dos currículos lattes.

O Brasil conta com dezesseis cursos de formação superior em Arquivologia, sendo que treze desses cursos se concentram em instituições de âmbito federal e três em âmbito estadual.

Neste contexto, mapeamos dezoito docentes que ministram disciplinas relacionadas às temáticas de descrição. Desses dezoito, onze possuem graduação em Arquivologia, seis são formados em História Licenciatura/ Bacharelado, três em Biblioteconomia, um em Administração de Empresas e um em Direito.

Verifica-se o predomínio da formação em Arquivologia pelos docentes que ministram disciplinas que tratam de temáticas relacionadas à descrição arquivística. Isso demonstra que, no decorrer de suas quase quatro décadas de existência, os cursos de Arquivologia brasileiros, já estão respondendo significativamente pela formação de docentes para a área. Porém, as formações em História e em Biblioteconomia, atuam como importantes subsídios para a formação desses docentes.

Ressalta-se, ainda, que entre esses docentes, dois têm dupla formação em Direito e Arquivologia, e Arquivologia e Licenciatura Plena em História, dois docentes têm formação de bacharel em História e também a Licenciatura Plena, constando no Lattes como duas formações. E ainda há um docente graduado em Biblioteconomia que esta cursando Arquivologia.

Quanto à formação em nível de mestrado, observamos que dos dezoito docentes identificados, dezessete apresentam essa formação, sendo que: nove fizeram em programas de Biblioteconomia/Ciência da Informação, três em História, dois em Patrimônio Cultural, um em Educação, um em Engenharia de produção e um docente esta com mestrado em andamento em História.

É possível observar, assim, o forte predomínio da formação nos programas de pós-graduação em mestrado em Ciência da Informação, o que demonstra que os programas de pós-graduação em Ciência da informação, assim como a ANCIB vêm revelando, ao longo dos anos, uma receptividade acerca das questões arquivísticas.

Na formação em nível de doutorado, fundamental para o docente, pois evidência seu caráter pleno de pesquisador, observou-se que oito docentes cursaram doutorado, sendo um em Difusão do conhecimento, dois em História, um em Educação, um em Letras, um em Memória Social, e dois em Ciência da Informação. Vale ressaltar que desses docentes, seis estão com doutorado em andamento, sendo que cinco deles estão cursando doutorado em programas de pós-graduação em Ciência da Informação, e um em Metodología y Investigación en Bibliotecas y Documentación, em universidade Espanhola.

Com a caraterização do panorama da formação dos docentes das disciplinas que tratam da temática descrição arquivística em cursos de Arquivologia do Brasil, tanto em graduação e em pós-graduação *stritu sensu,* cabe agora, analisar a produção científica desses docentes, que constitui [...] no conjunto de documentos gerados a partir dos resultados de pesquisas, publicados sob diferentes formas, e que representam a materialização do conhecimento gerado sobre um assunto de interesse de uma determinada comunidade científica. (DANUELLO, 2007, p.36-39),

Seguindo no mesmo contexto, analisou-se a produção científica desses docentes sendo consideradas as seguintes categorias: livro, capítulo de livro, artigo de periódico e trabalhos completos publicados em anais de eventos tendo em vista serem as publicações de maior impacto e visibilidade.

Para chegarmos a esses dados, identificamos em todas as produções disponíveis nos currículos *lattes* dos docentes, se nos títulos dos trabalhos apresentavam os seguintes termos: *descrição; representação da informação; representação arquivística; representação do conhecimento*, sendo que para esses termos procuramos identificar também a variação em língua inglesa e espanhola. Partindo daí, entramos em contato com os docentes para solicitar os trabalhos para identificação dos autores citados nas produções. Vale ressaltar, que alguns dados não foram apresentados, uma vez que a produção não eram atuais e os docentes não as possuíam mais em seus acervos.

A partir desse levantamento, analisaram-se as referências, desconsiderando-se as obras com autoria institucional e as autocitações. Dessa forma, chegou-se a um conjunto de 135 autores que receberam, no total, 207 citações, o que revela uma média de 1,6 citações por autor.

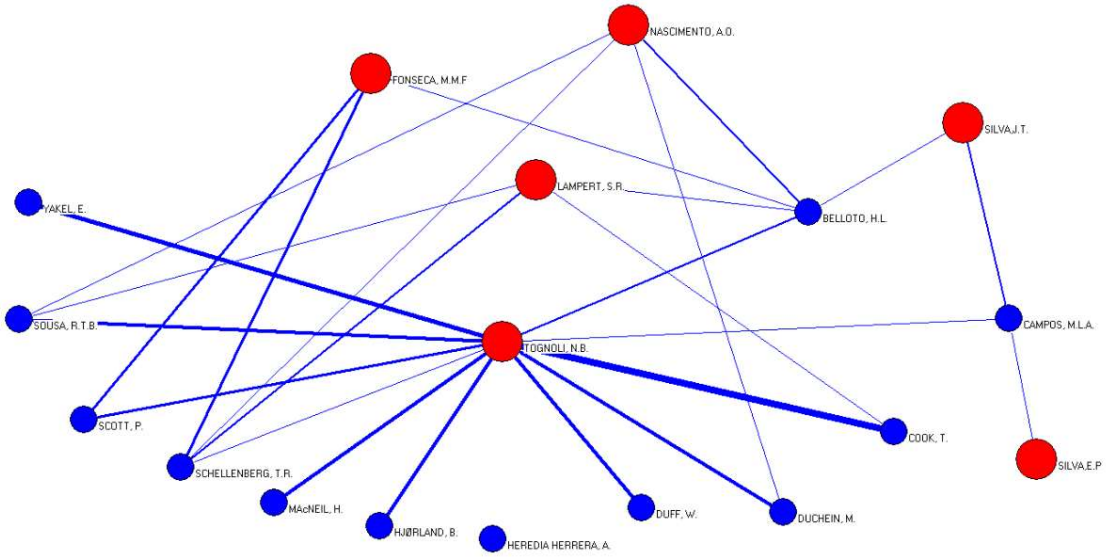
Para fins de análise desses referentes teóricos, aplicou-se a Lei de Elitismo de Price, que destaca que, [...] se k representa o número total de contribuintes numa disciplina, $\sqrt{k}$ representaria a elite da área estudada, assim como o número de contribuintes que gera a metade de todas as contribuições (URBIZAGÁSTEGUI ALVARADO, 2008, p.1).

Desse modo, chegou-se a um conjunto de 12 autores (9,6% do total) que receberam 4 ou mais citações cada, observando-se que esse universo responde por 29,8% do total das citações, a saber: Terry Cook, Heloísa Liberalli Bellotto, Theodore Roosevelt Schelenberg, Peter Scott, Renato Tarcísio Barbosa de Sousa, Michel

Duchein, Elizabeth Yakel, Maria Luiza Campos, Wendy Duff, Antonia Heredia Herrera, Birger Hjolrand e Heather McNeil.

Com base nesse conjunto de referentes teóricos, elaborou-se uma rede de citação por meio da utilização do software PAJEK, como se verifica a seguir:

Rede 1: Autores mais citados pelos docentes nas produções.



Fonte: Elaborado pelos autores

Sendo assim foi possível observar os autores mais citados por esses docentes em suas produções, nos quais chegamos aos seguintes dados: Terry Cook (Canadá) *nove* citações, Heloisa Liberalli Bellotto (Brasil) *sete*, Theodore Roosevelt Schellenberg (EUA) *sete*, Peter Scott (Austrália) *seis*, Renato Tarcísio Barbosa de Sousa (Brasil) *cinco*, Michel Duchein (França) *cinco*, Elizabeth Yakel (EUA) *cinco*, Maria Luísa de Almeida Campos (Brasil) *cinco*, Wendy Duff (Canadá) *quatro*, Antonia Heredia Herreira (Espanha) *quatro*, Birger Hjørland (Dinamarca) *quatro*, Heather MacNeil (Canadá) q*uatro* citações.

Considerações Finais

O presente estudo demonstrou, no âmbito da graduação, uma notável predominância da formação dos docentes em cursos de Arquivologia, o que

demonstra que os cursos estão conseguindo suprir as demandas por formação docente na área. No tocante aos cursos de pós-graduação, a predominância dos cursos em Ciência da Informação evidência, além, da receptividade da ANCIB às questões arquivísticas, uma aproximação da Arquivologia e de seus profissionais aos estudos da Ciência da Informação, fortalecendo a relação entre ambas às disciplinas.

Com relação à produção científica dos docentes, verificou-se que os cinco autores mais citados revelam efetivos referenciais teóricos na medida em que fornecem subsídios não apenas ao ensino, mas também à pesquisa na área, sendo eles: Terry Cook, Heloísa Bellotto, Antonia Heredia Herrera, Renato Barbosa de Sousa e Michel Duchein, o que proporciona um equilíbrio de influências teóricas, tendo em vista que foram identificados autores procedentes do Canadá, Espanha, Brasil e França, representantes de diferentes correntes teóricas.

Referências

BELLOTTO, H. L. **Arquivos Permanentes**: tratamento documental. 2a ed. Rio de Janeiro: FGV, 2004.

CRUZ MUNDET, J. R. **Manual de Archivistica**. 4. ed. Madrid: Fundação Germán Sanchez Ruipéres, 2001.

DANUELLO, J. C. **Produção científica docente em tratamento temático da informação no Brasil**: uma abordagem métrica como subsídio para a análise de domínio. Marília: Unesp, 2007. Dissertação (Mestrado) – Programa de Pós-Graduação em Ciência da Informação - Universidade Estadual Paulista (Unesp).

GUIMARÃES, J. A. C.; DANUELLO, J. C.; MENEZES, P. J. Ensino de tratamento temático da informação (T.T.I.) nos cursos de Biblioteconomia do Mercosul: uma análise de capacitação e produção científica docente com vistas ao delineamento de políticas integradas para área. In**: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO**, 5. 2003. Belo Horizonte. Anais. Belo Horizonte: Escola de Ciência da Informação da UFMG, 2003.

HEREDIA HERRERA, A. **Lenguaje y vocabulario archivísticos**: algo más que un diccionario. Sevilla, Junta de Andalucía, Consejería de Cultura, 2011.

MARQUES, A. A. C. **Os espaços e os diálogos da formação e configuração da Arquivística como disciplina no Brasil**. 2007. 298f. Dissertação (Mestrado) – Universidade de Brasília, Brasília, 2007.

MARQUES, A. A. C; RODRIGUES, G. M. Os cursos de arquivologia no Brasil: conquista de espaço acadêmico-institucional e delineamento de um campo científico.

In: **XV Congresso Brasileiro de Arquivologia**, 2008, Goiânia. Congresso Brasileiro de Arquivologia, 2008.

RODRIGUES, G. M. A representação da informação em Arquivística: uma abordagem a partir da perspectiva da norma internacional de descrição arquivística. In: RODRIGUES; G.M.; LOPES, I.L. (Org.). **Organização e representação do conhecimento na perspectiva da ciência da informação**. Brasília: Thesaurus, 2003, v. 2, p. 210-230.

TOGNOLI, N.B. A representação na arquivística contemporânea. **Revista Ibero-Americana de Ciência da Informação**, vol. 5, n. 2, 2012.

URBIZAGÁSTEGUI ALVARADO, R. Elitismo na literatura sobre a produtividade dos autores. **Ciência da Informação**, v.38, n.2, Maio/Ago. 2009. Disponível em: http://dx.doi.org/10.1590/S0100-19652009000200006.

Perspectivas da pesquisa em Organização do Conhecimento no Brasil

Knowledge Organization and the core of Information Science: tales of big data, computing clouds and social networks

Renato Rocha Souza
Getulio Vargas Foundation
renato.souza@fgv.br

Mauricio Barcellos Almeida
Federal University of Minas Gerais
mba@eci.ufmg.br

Renata Abrantes Baracho Porto
Federal University of Minas Gerais
renatabaracho@eci.ufmg

Introduction

Within all scientific fields, and even among those more recently established, one can say that Information Science (IS) is one of the most introspective fields with regards to research themes. The conceptual issues that underlie IS come up with very passionate arguments that their real research objects are sometimes blinded or faded into the background. Considering that seminal texts of the IS field are so broad – as seen in the works by Bush (1945) or Shanon and Weaver (1949) – and departing from such multifaceted concepts – again, as seen in Barlow (1994), Hofkirchner (1999) and Capurro (2003) – it is surprising that some authors are still able to find certain agreement about what IS should encompass under its umbrella (Borko, 1968; Zins, 2009). Indeed, the consolidation process of a relatively recent research field like IS, one find many obstacles if compared to the same processes in the so-called “hard-sciences”.

With respect to this relatively unique characteristic in the IS field, Wersig (1993) emphasizes the need for a pragmatic-evolutionary perspective along with attempts at theoretical construction. This should be done so that both empirical research objects and interfaces with other fields, work as a support system for the permanent construction and re-construction of the whole research field. Wersig (1993) describes the IS professional through an analogy with the “weaving bird”. It is a kind of bird that builds its nest by weaving elaborate and complex nets, which refers to the thematic connections, required in the ordinary work of IS professionals.

As IS is a dynamic research field, working on research objects somehow fluid and ubiquitous, we present here an opinion about how the paradigmatic and epistemological transformation occurs. There is a need for this field when considering the scenario of new technologies and new ways to handle information. We openly assume our intention of producing a certain provocation, as long as we believe that there is an on-going deflation of IS field as an autonomous research field. This deflation has occurred because of the increasing overlapping among research objects of IS and other fields. Such phenomenon has resulted in a substantial reduction of interest in the scope of IS, as well as the migration to other fields, whose research objects could be considered legitimate IS research objects.

The remaining part of the article is organized according to three scenarios: the deflation, the transformation and future perspectives.

Deflation

Borko (1968, p.3), in its remarkable definition that attempts to capture IS epistemology and praxis, asserts:

> Information Science is that discipline that investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation and utilization of information. This includes the investigation of information representations in both natural and artificial systems, the use of codes for efficient message transmission, and the study of information processing devices and techniques such as computers and their programming systems. It is an interdisciplinary science derived from and related to such fields as mathematics, logic, linguistics,

> psychology, computer technology, operations research, the graphic arts, communications, library science, management, and other similar fields.

Considering this definition, one might claim that virtually all research objects of the contemporary academic research would involve at least one component connected to IS. However, one can identify a bias, albeit broad, which can be read between the lines. The main concerns of IS are processes of information and knowledge representation, in addition to the manipulation of the records produced along these processes. The nature of these processes presupposes that information suffers successive abstractions, modeling and representations so that it can be organized, transmitted, codified and consumed. Buckland (1991) captures these processes and related objects by considering all of them as manifestations of the concept of information. Furthermore, it is worth emphasizing that information and knowledge representation are seminal subjects in the scope of IS.

Despite some discussion, there is a reasonable agreement about both the moment and the reasons for the birth of IS. On one hand, IS arises to deal with the issues that appeared from the increasing complexity in the management of informational collections. On the other hand, it fulfills the need for the creation and adaptation of methodologies – the praxis. Both aforementioned facts come from the fields of documentation and librarianship. However, the last few years of the 20th century were characterized by the fast expansion of information technology and by the acceleration of the processes of information transformation. These processes include the creation, representation, storage, organization, dissemination and consumption of information. This meant the issues from decades ago originally justifying the appearance of a so-called “Information Science”, are nowadays exponentially felt.

Networks, mobile devices, tablets and other gadgets, as well as digital libraries and emergent developments like wearable computing, have continuously changed contexts, shorten cycles, and reinvented material supports. These events have redefined the relation between people and information records. There is a progressive “undocking” process, chagig deeply the usual connections between information and its supports for recording, exchanging and consuming. Organization of the large mass of data needs new and creative solutions, suggesting the real need of a science to deal with information, namely IS.

This “technological ecology” could favor a new inspiration to the IS field, by multiplying the informational issues and the consequent demand for new solutions. But what has really occurred is the gradual migration of genuine IS research objects to other fields. To illustrate such thematic migration, one can mention socio-technical artifacts for information retrieval (like digital libraries), instruments for knowledge representation (like ontologies), techniques for domain modeling, information systems studies in their several contexts of use, among others. All of these examples of research objects have been actively studied by other sciences in fields of Linguistics, Administration, Computer Science and Information Systems.

On the other hand, themes like “social responsibility”, “inter and transdisciplinarity”, “information and work”, “information society”, and so forth (ANCIB, 2009), although relevant and legitimate themes for a social and applied science, have directing the field to focus on aspects related to the social issues coming from Sociology, Anthropology, History, among others. The rise of research in themes related to management, as Knowledge Management and Competitive Intelligence make IS close to fields of Administration and of some subfields of Engineering. All these examples show borderline overlapping that contribute to dissolving IS frontiers and to the lack of its identity.

With regard to the processes of information and knowledge representation, which are strongly connected to the origins of the IS field, one can expect that it become a mere and passive spectator in a highly dynamic environment. Even though IS is a dominant field when the subject is the study and construction of instruments for information organization, like indexing languages and information systems interfaces, these themes are nowadays associated with “technical sciences”. Indeed, these sciences have incorporated these instruments into their constructs. Borko (1968) analyses the research objects mentioned in the “Current Research and Development in Scientific Documentation” and lists nine categories, namely:

- Information need and uses
- Document creation and copying
- Language analysis
- Translation
- Abstracting, classification, coding and indexing
- System design

- Analysis and evaluation
- Pattern recognition
- Adaptive systems

Among these objects, which are very current topics, only 1, 2 and part of 5 continue to be recognized as part of the core of IS. Linguistic analysis has been properly developed in fields like Computational Linguistics, Corpus Linguistics, Computer Science and even Applied Mathematics. Topics 3 and 4, as well as part of 5 and 7, have been developed under the label of Natural Language Processing. Computer Science and Engineering encompass topic 6, namely, Systems Design. Finally, the growing field of Artificial Intelligence has developed the topics 8 and 9.

In many universities around the world, the "schools of information science" have been transformed into "schools for information studies", where special groups gather the required profiles able to exert the socio-technical interdisciplinary skills. In the United States, an example of this sort of discussion is the so-called *i-school´s movement*, in which several renowned universities have adopted new techniques, attitudes and even new curricula in schools and departments (Detlefsen, 2008). In addition, the conferences of the International Society for Knowledge Organization (ISKO 2014) also denote a clear direction that seeks a broad coverage, while maintaining the theoretical focus. The incorporation of themes that would suggest a social bias, as the *folksonomies*, occur in the scope of the knowledge organization systems, a very dear and traditional research subject within IS.

We conclude this first scenario emphasizing the need to take care of IS field, so that it does not become a mere niche among other fields. Some hope that information professionals, instead of being the Wersig´s weaving bird, become a sort of "remora"[16], which feeds on the thematic leftovers of topics that other fields develop.

Transformation

There are surveys and research available in the literature, which characterize IS under countless different aspects, for example Capurro (1991) and Whittaker (2011), to mention a few. The scenarios presented here in this section do not

[16] Parasite fishes that have on the head a sucking disk with which they attach themselves to sharks and other fishes so that they can take rests of food.

correspond to an exhaustive literature review. Instead, we chose just to approach the effects of some new developments in IS, namely, Big Data, Cloud Computing and Social Networks. We believe that those developments may have caused impacts and paradigmatic changes in IS research.

0 Social Networks

Social networks are structures that gather actors – individuals and institutions – through links (Easley et al, 2010), which can be reified as social or technological arrangements. The representation and study of those networks include, but are not limited to, fields of Anthropology, Biology, Communication, Economy, Geography, Information Science, Psychology, Sociology and Socio-Linguistics.

Here, we only approach the social networks constructed from socio-technical arrangements, as is the case of some available on the web, for example, networks for relationships like Twitter[17] and Facebook[18]. In its early days, Twitter was discarded because it was not considered a relevant information source, but in 2010 its economic messages of 140 characters started to be collected[19] by the United States Library of Congress. In 2014, more than 500 million messages were exchanged in Twitter each day[20]. Similarly, in 2014 Facebook had around 1.1 billion users exchanging an amazing volume of information[21]. Only these two instruments completely surpassed the information currently disseminated in printed or digital media. The main feature of this sort of information is the fluidity, dynamics and thematic ephemerality, which are explored in activities of environmental monitoring, sentiment analysis (Liu, 2012) and even the recording of human history.

It is worth mentioning another phenomenon, Wikipedia, one of the most remarkable crowd sourced collective developments of our age. Wikipedia is a free and collaborative multilingual encyclopedia, which makes it possible to copy, change and extend any type of information. It is distinguished from blogs because it does not use the traditional concept of authorship. Anyone with Internet access is able to modify any

[17] https://twitter.com/

[18] https://pt-br.facebook.com/

[19] http://www.businessinsider.com/library-of-congress-is-archiving-all-of-americas-tweets-2013-1

[20] http://www.internetlivestats.com/twitter-statistics/

[21] http://www.statisticbrain.com/facebook-statistics/

Wikipedia article, and each reader is a potential collaborator. In addition, Wikipedia goes beyond the results obtained by search engines, as the response to a query is a single text.

It is still worth remembering that social networks and their related phenomena point out to the ephemeral character of both the informational records and the instruments required to organize these records. Indeed, these collective constructions allow one to create products of unimaginable extent, which will require new approaches and studies (Surowiecki, 2004).

1 Big data

The explosion of information, which is sometimes called information or data deluge, is more than a large increase in the number of records. The term “big data” has become somewhat commonplace: it is used to refer to any mass of data, which cannot be processed without specific computational structures. In addition to the large volume, some definitions assigned to big data features like variety, velocity, and veracity (Laney, 2012). Indeed, since the 1990s large volumes of data have been manipulated with the aim of producing simulations and predictions. However, the textual non-structured information has been increasing a lot since the 2000s. This phenomenon has had a strong impact on the academic community and in research, most notably in fields like IS. The professionals using traditional techniques of subject analysis, classification and categorization have been trying to adapt their expertise, both to the gigantic collections, and to the speed of production of summaries and indexes (Magnusson; Vanharanta, 2003; Nodus Labs, 2012).

In the scope of academic communities, works of literature review, which are essential for research, are increasingly complex and in some cases impractical. Often a researcher has to cut arbitrary and explicit excerpts from the literature, meaning that the quantity of publications and available sources about specific subjects has become intractable. In other scientific cases, it is now a common research mode called “data driven research” (Moe-Behrens, 2012). According to this mode, one should perform several tests of hypotheses in large masses of data as an exploratory preamble. When promising possibilities are found from the data analysis, it would be possible to propose qualitative and quantitative enhancements though interactive changes of focus.

2 Cloud computing

Since we have already briefly described both the social phenomenon materialized through networks, and explosion of information illustrated through the hype around the big data, it is worth mentioning the importance of clouding computing. This phenomenon adds one more abstract level to the process of dematerialization of information records. Some time ago, still without confidence, people changed from physical supports to digital ones like tablets and mobile devices. But now, the cloud computing separates the users from those devices that store their data. Ubiquitous and often freely available storage services (DropBox[22], Google Drive[23], iCloud[24], to mention a few) are part of the lives of millions of people. These people seen unconcerned with the potential lack or theft of both photos and private data, even considering the current quantity of security incidents.

So, one can notice that: (a) social networks have disrupted the geographical limits of connectivity; (b) the new explosion of information has eliminated the possibility for an individualized treatment of information records; and (c) cloud computing has broken up the tangibility of records. So, in this complex context, an important issue begs for a proper answer: what will be the research agenda for IS in the 21st century?

- **Future Perspectives**

What would be the role of IS in a context in which it seems to have lost primacy as a research field, and even centrality, in the thematic of information and knowledge representation? Resuming our initial provocation in the aforementioned scenario of IS deflation, what will be the possible arrangements capable of avoiding such deflation? Revisiting the definition of IS by Borko (1968), as “an interdisciplinary science that investigates the properties and behavior of information, the forces that govern the flow and use of information and the techniques of processing”, one can conclude that this program has become too broad for the IS field. Indeed, one can even conclude that such a program is too broad for any research field.

Yet, the already cited proposal of Wersig (1993) has become more and more present. It recognizes the new role of knowledge in contemporary society, and

[22] https://www.dropbox.com/

[23] https://drive.google.com/

[24] https://www.icloud.com/

advocates the adoption of pragmatism to deal with problems in IS. Thus, in this scenario, epistemological approaches should receive less emphasis. Wersig´s third model for IS suggests, on one hand, a theoretical structure that reduces the attempts of formulating general laws. On the other hand, it suggests the creation of more action strategies from an approach based on the interlacement of scientific concepts. Then, it would weave a proto-net of basic concepts in IS, from which other people or groups could continuously find and integrate new concepts. In this sense, this conceptual net would become more inclusive and strong, in addition to having augmented it´s scientific character.

Examining this from another perspective, on one hand, interdisciplinary studies may weak IS as a research field. But, on the other hand, this same interdisciplinary studies may bring the possibility, and even the prerogative, of mediation among disciplinary dialogues. Such interdisciplinary essence fosters the professional and the researcher of IS to navigate in new theoretical spaces, to adapt to the technological contexts and to continuously reinvent itself.

Curriculum Vitae

Renato Rocha Souza holds a PhD in Information Science and currently acts as Adjunct Professor in the School of Applied Mathematics at the *Getulio Vargas* Foundation, Brazil.

Mauricio Barcellos Almeida holds a PhD in Information Science and currently acts as Associate Professor in the Department of Information Theory and Management at the Federal University of *Minas Gerais*, Brazil.

Renata Baracho Porto holds a PhD in Information Science and currently acts as Adjunct Professor in the Department of Information Theory and Management at the Federal University *of Minas Gerais*, Brazil.

References

ANCIB. (2009). History of the Brazilian Conferences of Research in Information Science. Retrieved July 31, 2011 from http://www.ancib.org.br/enancib/historico-do-enancib/

Borko, H. (1968). Information science: what is this? American Documentation, v.19, p.03-05.

Bush, V. (1945). As we may think, in the atlantic monthly. Retrieved August 22, 2010 from http://ccat.sas.upenn.edu/ jod/texts/vannevar.bush.html

Capurro, R. (2003). The concept of information. Annual Review of Information Science and Technology (ARIST), v.37, n.8, p.343-411.

Capurro, R. (1991). What is Information Science for? a philosophical reflection. In: Vakkari, P.; Cronin, B. (Ed.). Conceptions of Library and Information Science. Proceedings... London: Taylor Graham, p.82-96.

Detlefsen, E. G. What´s is a name? the I-Schools project. Medical Library Association News, v.28, n.3, p.1-17, 2008.

Easley, D.; Kleinberg, J. (2010). Networks, Crowds, and Markets: reasoning about a highly connected world. Cambridge: University Press.

Hofkirchner, W. (1999). The quest for a unified theory of information. In: Hofkirchner, W. (Ed.). Proceedings of the 2nd Intern. Conf. on the Foundations of Information Science. Amsterdam: Gordon & Breach, p. 9-30.

Ingwersen, P.; Järvelin, K.. (2005). The turn: integration of information seeking and retrieval in context. Netherlands: Springer, v. 18.

International Society For Knowledge Organization. ISKO Chapters. Retrieved March 29, 2014 from http://www.isko.org/chapters.html

Laney, D. (2012). The Importance of 'Big Data': a definition. Retrieved March 22, 2014 from https://www.gartner.com/doc/2057415/importance-big-data-definition
Liu, B. (2012). Sentiment Analysis and Opinion Mining, Chicago: Morgan & Claypool.
Magnusson, C.; Vanharanta, H. (2003). Visualizing sequences of texts using collocational networks. In: 3rd Intern. Conf. Leipzig. v.2734, p. 276-283.

Moe-Behrens, G. (2012). From hypothesis to data driven research. Retrieved April 2, 2014 from http://www.sciencetogrok.com/2012/12/from-hypothesis-to-data-driven-research.html

Nodus Labs. (2012). Textexture: visualize any text as a network. Retrieved July 11, 2013 from http://textexture.com/

Osterberg, G. Update on the Twitter archive at the Library of Congress. Retrieved July 3, 2014 from http://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress

Shannon, C.E.; Weaver, W. (1949). The Mathematical theory of communication. Urbana: University of Illinois Press.

Surowiecki, J. (2004). The Wisdom of Crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. New York: Anchor.

Vakkari, P.; Cronin, B. (1992). Conceptions of Library and Information Science. Historical, empirical and theoretical perspectives. London: Taylor Graham.

Wersig, G. (1993). Information science: the study of postmodern knowledge usage. Information Processing & Management, v.29, n.2, p.229-239.

Whittaker, S. (2011). Personal information management: from information consumption to curation. Annual Review of Information Science and Technology. v.45, n.1, p. 1-62, 2011.

Zins, C. (2009). Knowledge map of information science: issues, principles, implications. Retrieved September 30, 2012 from http://www.success.co.il/is/index.html

Organização do conhecimento: pesquisa e desenvolvimento

Gercina A. B. O. LIMA
Universidade Federal de Minas Gerais
limagercina@gmail.com

1 Introdução

O presente trabalho tem por objetivo verificar as tendências de pesquisa na área de Organização do Conhecimento (OC) no Brasil. Porém sentiu-se necessidade de, primeiramente, esclarecer essas características e sobreposições existentes entre as áreas, porque as temáticas de pesquisas estudadas na literatura poderão ter abordagens vindas tanto das áreas da Classificação, quanto da Organização do Conhecimento e da Biblioteconomia e Ciência da Informação (BCI).

A área de organização do conhecimento tem suas raízes históricas na base teórica da classificação. A Classificação pode ser vista como uma área de conhecimento, a qual classifica o conhecimento filosófico, que se inicia com o modelo categorial proposto por Aristóteles; e como classificação bibliográfica que possui processos e instrumentos para representar conteúdo temático de documentos e aumentar a eficácia na recuperação de informações. As duas abordagens da classificação procuram ligar teoria e prática nos estudos e pesquisas realizadas dentro da área da Biblioteconomia e Ciência da Informação (BCI). Tal como a Classificação, a Organização do Conhecimento (OC) é considerada, também, uma área do conhecimento e uma atividade, e ambas são consideradas áreas nucleares no âmbito da BCI.

De acordo com Dalhberg (2006), a Organização do Conhecimento como área de conhecimento tem sua história ligada à criação da *International Society of Knowledge Organization* (ISKO), em 22 de julho de 1989. Dalhberg (1993) esclarece

que o termo *organização do conhecimento* já tinha sido utilizado nos titulos das obras de Henry Evelyn Bliss, *The Organization of Knowledge Organization and the system of the science*, e *The Organization of Knowlodge Organization in Libraries*, publicados, respectivamente, em 1926 e 1936. Essa expressão foi utilizada, também, no titulo da tese de Dagobert Soergel, intitulada *Organization of knowledge and documentation* (1971), defendida na Alemanha, e na tese da propria autora, que foi publicada sob o título *Foundation of Universal Organization of Knowledge.*

O periódico internacional *Knowlodge Organization* (KO), publicado sob a responsabilidade da ISKO, começou a ser publicado em 1993, substituindo o periódico *International Classification*, que foi publicado de 1974 a 1992. Dalhberg (1993) explica que os membros da ISKO decidiram que o conceito “classificação” seria, a partir daí, interpretado como um método de classificação, que agrupa objetos semelhantes. Porém, o periodico *Knowledge Organization* traz como missão a divulgação de trabalhos que representem contribuições para o desenvolvimento de novos conhecimentos sobre a teoria do conceito, classificação, indexação e representação do conhecimento. Nota-se, portanto, que as áreas Classificação e Organização do Conhecimento são muito próximas, tanto no embasamento teórico, quanto em suas atividades práticas.

McIlwaine e Williamson (1992) afirmam que algumas áreas que são pesquisadas não são tão novas como se supõe, pois a teoria e princípios da Classificação serão sempre fundamentais para todos os aspectos da organização do conhecimento.

Hagar Gomes (2009), em seu trabalho “Tendências da pesquisa em organização do conhecimento”, aponta que:

> Percebe-se, então, que a OC é uma área de múltiplas aplicações. As pesquisas especificamente na área apontam para, no mínimo, dois aspectos importantes: de um lado, a abordagem teórica-filosófica voltada para questões epistemológicas ao analisar um domínio juntamente com as aplicações na organização de sistemas/tabelas de classificação e, de outro lado, os métodos empregados na elaboração propriamente dita dos sistemas, atividades que alguns autores consideram suplementar, mas que tem suas bases teóricas já sedimentadas (GOMES, 2009, p.63).

A autora acrescenta, ainda, que a área da organização do conhecimento é muito próxima da Biblioteconomia e Ciência da Informação (BCI), dado que aponta

para a ideia de que a própria literatura ainda confunde as bases teóricas de cada área, e não tem definido claramente quais atividades estão inseridas somente em uma delas, ou se estão todas relacionadas. Talvez seja por isso que Smiraglia (2005, 2006), editor do periódico *Knowlodge Organization*, demostra sua preocupação quando aponta "*[...] what precisaly KO is [...] although many have written about it."*, (2005. p..139) demonstrando, assim, que ainda não existe um consenso geral sobre a definição da área, voltando à discussão inicial de quando KO era originalmente chamada de *Classificação*.

Outra observação que deve ser feita se refere à utilização dos termos *organização do conhecimento* (OC) e *organização da informação* (OI), que têm sido usados em diferentes contextos, por diversos autores da área. Bräscher e Café (2010) apresentam um estudo sobre a utilização desses termos dentro da área da Ciência da Informação (CI). As autoras concluem que tal maneira de se utilizar os termos demonstra a falta de delimitação do conceito em questão, e afirmam que:

> Por vezes, o termo organização do conhecimento é utilizado no sentido de organização da informação, e vice-versa e, em determinadas situações, empregam-se os termos conjuntamente, organização da informação e do conhecimento. (BRACHER; CAFÉ, 2010, p.85).

Na visão das autoras, existem dois tipos de processos de organização: a organização da informação que tem o objetivo de descrever as características físicas e conceituais dos objetos informacionais, próximos às atividades realizadas em bibliotecas e centros de documentação; e a organização do conhecimento que visa à criação de modelos que representam as unidades de conhecimento de um determinado domínio, de acordo com suas características. Assim, elas apontam que,

> Esses dois processos (OI e OC) produzem, consequentemente, dois tipos distintos de representação: a representação da informação, compreendida como o conjunto de atributos que representa determinado objeto informacional e que é obtido pelos processos de descrição física e de conteúdo, e a representação do conhecimento, que se constitui numa estrutura conceitual que representa modelos de mundo. (BRACHER; CAFÉ, 2010, p.93).

Pode-se dizer, então, que esses dois processos se complementam, pois, enquanto na área de OI ocorre o processo de representação do conteúdo de um documento especifico, o qual tem o conhecimento registrado, para que ele seja

posteriormente recuperado pelo usuário em um Sistema de recuperação da Informação (SRI), a OC organiza o conhecimento a partir da modelagem de domínios, representando sua estrutura com base na descrição de conceitos e dos relacionamentos semânticos entre eles. As estruturas geralmente são criadas com o auxílio de especialistas da área, que validam o conhecimento ali organizado a partir de suas expertises.

Nesta pesquisa foram considerados os trabalhos que tratam das tendências de pesquisa na área de Organização do Conhecimento, sem, no entanto, diferenciá-los dessas outras áreas, pois, em muitos títulos dos trabalhos publicados, não existe uma padronização, conforme já foi relatado na literatura.

Para mapear estas tendências, propôs-se primeiramente, fazer uma breve revisão sobre a pesquisa na área de OC, buscando-se artigos de periódicos e anais de eventos nacionais e internacionais, sem a pretensão de uma busca exaustiva. Portanto, salienta-se que o resultado da pesquisa bibliográfica teve seu recorte temporal entre os anos de 1993 e 2015.

Para o levantamento bibliográfico foram utilizadas as seguintes expressões, em português e em inglês, na estratégia de busca: *tendências de pesquisa em organização do conhecimento*, *pesquisa em organização da informação*, *organização e representação da informação no Brasil*, *ensino e pesquisa em ciência da informação*, *current trends in Knowlede Organization, Knowledge Organization research, trends in classification research*.

2 Tendências da pesquisa em organização do conhecimento através do tempo: uma sumarização

Qualquer pesquisa em um determinado campo de conhecimento avança a partir de perguntas feitas considerando-se seus antecedentes históricos e a experiência intelectual já desenvolvida pelos pesquisadores da área. Essas abordagens aumentam nossa compreensão sobre o que já foi pesquisado e desenvolvido, e nos auxiliam a entender as atividades no presente, permitindo que sejam levantadas as lacunas que necessitam serem estudadas e perguntas de pesquisas pertinentes para o seu desenvolvimento.

Para a apresentação dos trabalhos selecionados que apontam tendências de pesquisa na área de organização do conhecimento, optou-se por descrever de forma

sumarizada, primeiramente, os artigos estrangeiros sobre o tema, e posteriormente, os artigos brasileiros, considerando a ordem cronológica de exposição, para, em seguida apresentar os temas de pesquisa que têm sido apontados como necessários para o desenvolvimento da área no Brasil, atualmente.

Em 1981, Elaine Svenonius já apresentava preocupação em relação à pesquisa na, então, chamada área de Controle Bibliográfico. Em seu artigo intitulado “Directions for reseach in indexing, classification, and cataloging”, a autora aponta para a necessidade de se fazer pesquisa na área para resolver problemas na organização do conhecimento/informação para melhorar sua recuperação. Em seu trabalho, a autora sugere os seguintes estudos: (1) dificuldade de medir a performance da recuperação da informação nos sistemas de recuperação da informação (SRI), porque esta depende diretamente relevância, e as medidas de precisão e revocação não têm auxiliado, porque nem sempre, em suas aplicações, levam em conta as necessidades dos usuários, (2) necessidade de melhorar os relacionamentos semânticos nas linguagens de indexação, (3) necessidade de melhorar a indexação automática desenvolvendo técnicas de extração mais eficientes – neste caso, a autora sugere a contribuição da linguística computacional e a utilização da inteligência artificial; (4) necessidade de estudos com abordagem mais teórica para avaliar os sistemas de classificação bibliográfica, para melhorar a adaptabilidade e a hospitalidade, principalmente para domínios específicos, (5) pesquisas sobre a compatibilidade, a conversão e a tradução dos vocabulários controlados para solucionar o problema da interoperabilidade dos sistemas.

O artigo “Current trends in knwledge organization” foi publicado por Ingetraut Dahlberg, em 1993, e, nele, a autora apresenta o estado da arte sobre os avanços e tendências da pesquisa em Organização do Conhecimento. Após apresentar uma introdução histórica da área, a Dahlberg apresenta um estudo sistematizado, tendo por base a análise da literatura da área, com o intuito de verificar quais seriam as tendências de pesquisa naquele momento. Os resultados apontaram para a necessidade de pesquisa em: (1) automação da organização do conhecimento, incluindo, aqui, a automação dos sistemas de classificação universais; (2) estudos sobre a criação de um tesauro universal, nos moldes dos dicionários linguísticos ou com base na estrutura facetada; (3) compatibilidade entre os sistemas de classificação bibliográfica com os tesauros já existentes, principalmente para atender a domínios específicos do conhecimento; (4) desenvolvimento de estruturas para os sistemas de

classificação bibliográfica e tesauros por meio de estudos sobre as categorias e seus relacionamentos; (5) estudos sobre a utilização da programação orientada a objetos; (6) melhoria na qualidade dos processos de indexação e catalogação de assunto; (7) necessidade de aproximação dos profissionais da área de organização do conhecimento, como os terminológos, para que possam utilizar as metodologias da terminologia na definição dos conceitos.

Hjorland e Albrechtsen (1999) publicam um estudo intitulado “An analysis of some trends in classification research”, no qual analisam as três questões formuladas por Clare Begthol (1998), então presidente da ISKO. Para a autora, essas questões são nucleares para a área da organização do conhecimento:

> 1) the academic disciplines as the main structural principle; 2) the fiction/nonfiction distintion as one secondary structural principle; and 3) information retrieval techniques that call into a question whether a whole document (e.g; book, article) is the most appropriate unity of analysis in online retrieval systems. (BEGTHOL, 1998, p.1).

Os autores fazem uma análise, do ponto de vista epistemológico, dessas três questões e tecem a seguintes observações: (1) as disciplinas acadêmicas fornecem princípios para a estrutura central ou fundamental no sistema de divisão dos esquemas de classificação; (2) avaliação de como a classificação do conhecimento pode representar algumas áreas especificas do conhecimento, sem perder a garantia cultural e o viés epistemológico, por isso os sistemas de classificação bibliográfica devem refletir as áreas interdisciplinares e as relações entre as áreas do conhecimento; (3) em relação ao conhecimento ser dividido em ficção e não ficção, Hjorland e Albrechtsen (1999) reconhecem que esta questão não deve ser considerada fundamental, mas que, em alguns tipos de conhecimento, tais como os da área de humanidades, como a Psicologia, por exemplo, devem ser representados sob diferentes aspectos; (4) a indexação de capítulos e artigos já é uma pratica comum dentro de um sistema de recuperação da informação, portanto, os usuários precisam de acesso aos documentos seja através das partes, ou seja, de capítulos ou de periódicos, seja do documento no todo, considerando-se que as unidades de analises estão relacionadas ao conteúdo, e não à estrutura do documento; (5) necessidade estudos para melhorar os pontos de acesso de assunto nas bases de dados bibliográficas ou de textos completos, para otimizar a recuperação da informação.

Também no ano de 1999, McIlwaine e Williamson apresentam o artigo intitulado “International trends in subject analysis reseach”, no qual descrevem a pesquisa realizada sobre o tema pesquisa análise de assunto, no período de 10 anos: de 1988 a 1998. As autoras concentram a pesquisas em três áreas: (1) sistemas de classificação bibliográfica universal, (2) desenvolvimento e *design* de tesauros, e (3) esforços para organização da Internet. As autoras relatam que: (1) o progresso sobre a conversão dos três sistemas mais usados, CDD, CDU e LCC, em formato legível por computador foi um grande avanço para os sistemas de classificação universal; (2) esses sistemas podem ser usados na melhoria da recuperação da informação na Internet. Em relação aos tesauros, o principal tópico de pesquisa relatado foi a aplicação da análise facetada na construção de tesauros; (3) o tesauro começou a ser pesquisado como uma ferramenta de organização e navegação a ser utilizada na Internet; (4) havia a necessidade de criação de um tesauro multilíngue, não com a intenção de traduzir os termos, mas realizar pesquisas que possibilitem a compatibilização entre tesauros e identificação de metodologias para que eles possam ser acessados e manipulados *online*; (5) o processo classificatório já tem sido utilizado em alguns *sites* da Internet, mas de uma maneira pouco elaborada. Porém, carece de estudos sobre as interfaces interativas, ordem de citação, mineração de dados, tesauros em formato hipertextual e agentes inteligentes.

Em 2003, McIlwaine apresenta o artigo “Trends in Knowledge organization research”, concluindo que: (1) a interoperabilidade entre os sistemas é um dos tópicos mais importantes para melhorar o acesso à Internet através da abordagem da inteligência artificial, utilizando métodos da linguística, como, por exemplo, o processamento em linguagem natural; (2) a maioria das pesquisas acontece em um ambiente acadêmico, com recursos governamentais, geralmente disponibilizados por um curto período de tempo, o que dificulta a manutenção e a finalização das pesquisas, na maioria das vezes; (3) para resolver o problema em relação à língua, a autora sugere a criação de vocabulários que estejam ligados com um sistema verbal, de preferência multilíngue, elaborado com a ajuda de profissionais da linguística, da inteligência artificial, para resolver o problema da interoperabilidade, do multilinguismo, e da interface interativa para o usuário.

Lopèz-Heurtas, em 2008, apresenta seu artigo “Some currents reseach questions in the field of knowledge organization” uma exaustiva revisão de literatura e aponta a necessidade de pesquisas: (1) melhorar a qualidade na organização e

representação do conhecimento e recuperação da informação no ambiente da Internet; (2) o multilinguismo na organização do conhecimento; (3) estudos sobre a integração das diferenças socioculturais na organização do conhecimento para melhorar a qualidade dos sistemas de organização do conhecimento (SOC); (4) melhor representação e integração dos setores considerados marginalizados e minoritários nos sistemas de organização do conhecimento; (5) integração de conhecimento na construção de SOC para facilitar a integração de diversas estruturas na *web*; (6) a organização e a representação do conhecimento multidimensional em SOCs; (7) o uso da análise de domínio para o conhecimento e gestão em espaços interdisciplinares e transdisciplinares, através da abordagem bibliométrica, e terminológica da análise de domínio; (8) criação de ontologias e sistemas interdisciplinares;

O artigo “Knowledge organization trends in library and information studies: a preliminar comparison of the pre- and post-web eras”, de Saumure e Shiri publicado em 2008, cobre o período de 1966 a 2006. Na análise qualitativa realizada pelos autores, foram identificadas as seguintes temáticas que representam a tendência de pesquisa nas fases pré-*web* e pós-*web*, com uma amostra total de 219 artigos analisados: (1) organização de informações corporativas ou de negócios; (2) organização do conhecimento assistida por máquina, (3) bibliotecários como organizadores de conhecimento na *web*, (4) interoperabilidade, (5) catalogação e classificação, (6) classificação de informações na *web*, (7) preservação digital e as bibliotecas digitais, (8) aplicações de metadados e usos, (9) cognição, (10) educação, (11) indexação e resumo, (12) tesauros (iniciativas).

Nos gráficos a seguir, são apresentadas as temáticas dos 31 artigos selecionados para a representação da fase pré-web (GRÁFICO 1) e os 188 artigos selecionados na fase pós-*web* (GRAFICO 2).

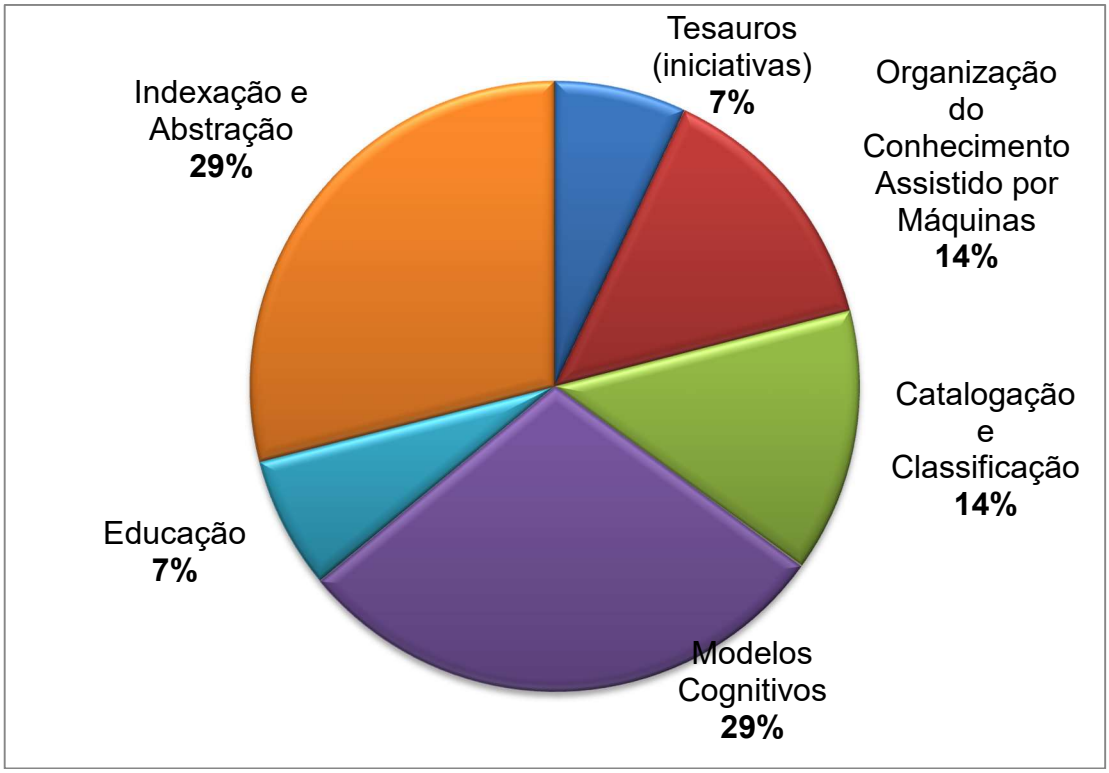


Gráfico 1 – Representação numérica de temas pré-*web*

Fonte: traduzido de Saumure e Shiri (2008, p.654).

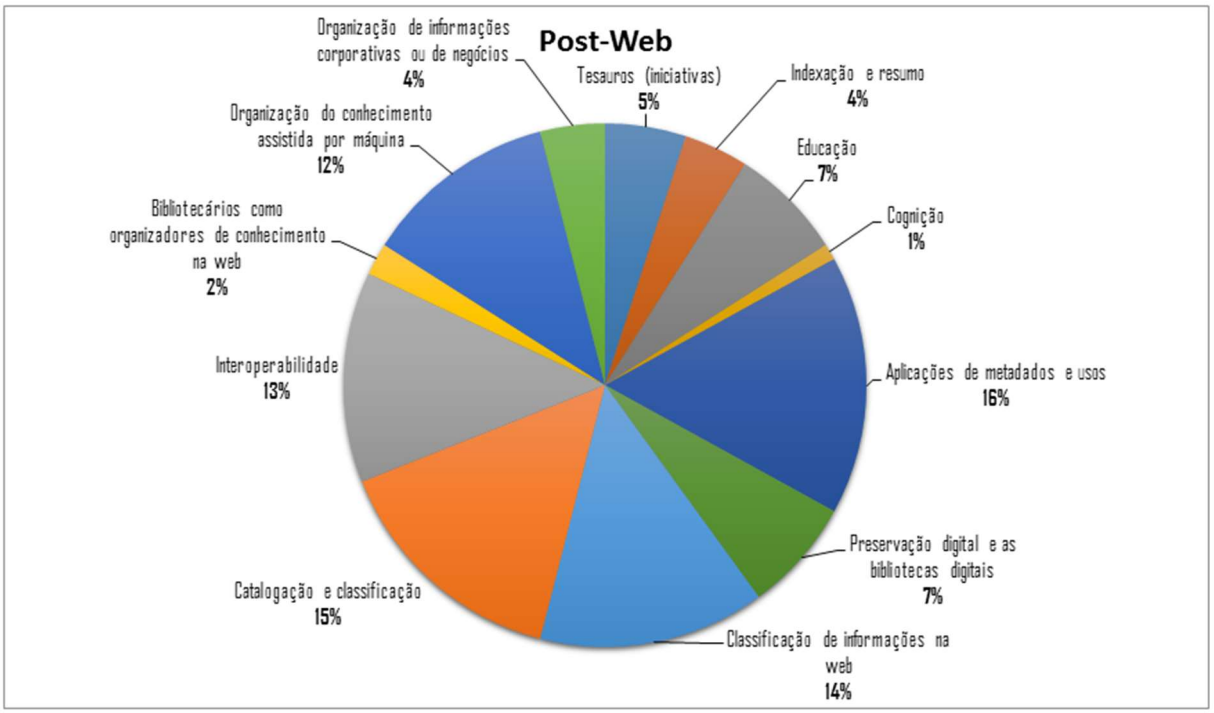

Post-Web
Organização de informações corporativas ou de negócios 4%
Tesauros (iniciativas) 5%
Indexação e resumo 4%
Organização do conhecimento assistida por máquina 12%
Educação 7%
Bibliotecários como organizadores de conhecimento na web 2%
Cognição 1%
Interoperabilidade 13%
Aplicações de metadados e usos 16%
Preservação digital e as bibliotecas digitais 7%
Catalogação e classificação 15%
Classificação de informações na web 14%

Gráfico 2 – Representação numérica de temas pós-*web*

Fonte: traduzido de Saumure e Shiri (2008, p.654).

Em 2010, Ibekwe-sanjuan e Sanjuan publicam o artigo “Knowledge organization research in the last two decades: 1988-2008”, no qual os autores fazem uma revisão de literatura citando as pesquisas mencionadas no presente trabalho: Hjorland e Albrechsten (1999), Mcilwaine (2003), Lopez-huertas (2008) e Saumure e Shiri (2008). Os autores relatam que a maioria destes estudos não deixou explícita a metodologia utilizada, de forma que a leitura e interpretação dos dados foi realizada com base no conhecimento dos autores supracitados, que leram e analisaram as pesquisas fazendo inferências e tirando suas conclusões. Para realizar esse estudo, os autores utilizam o método bibliométrico e a análise de dados para identificar os tópicos em organização do conhecimento em 31 periódicos internacionais pesquisado na *Web of Science* (WoS), separados em duas décadas. Como resultado da primeira década (1988-1997), os autores obtiveram os seguintes tópicos de pesquisa: (1) conhecimento, (2) classificação, (3) organização do conhecimento, (4) vocabulário controlado, (5) *design* de base de dados bibliográficos, (6) indexação, (7) construção e uso de tesauros, (8) informação e análise de texto, (9) informação e documentação, (10) Ciência da Informação, (11) organização e representação do conhecimento e (11) Sistemas de classificação bibliográfica. Na segunda década (1998-2008), foram mapeados os seguintes tópicos: (1) classificação, (2) informação, (3) conhecimento, (4) organização do conhecimento, com seus diferentes aspectos (transferência, perspectivas, domínio, integração, administração, pesquisa, mapas, revocação, representação, redes, e literatura), (5) ciência da informação, (6) fundamentos epistemológicos, (7) base de dados terminológica, (8) recuperação da informação, (9) *metadata*, (10) vocabulário controlado especificamente para tratar a temática sobre homossexuais, (11) *Web*. Deve-se ressaltar que estes três últimos tópicos não estavam presentes na primeira década.

O último artigo internacional a ser apresentado é o de Beak et al (2013), intitulado “International comparative domain analysis in knowledge organization research topics in four countries - Brazil, South Korea, Spain and the United States”. Para realizar este trabalho, cada autor escolheu dois periódicos de seu país dentro da área da Biblioteconomia e Ciência da Informação (BCI). Os autores apresentam análises estatísticas que comparam quais termos se sobrepõem entre os quatro países, quais termos aparecem somente em um dos países, e, também, fazem uma análise sobre as ocorrências entre cada um dos países em relação a cada um dos

outros. Para ilustrar a ocorrência dos termos em cada país, apresenta-se a seguir o quadro 1, extraída do artigo.

Gráfico 2 – Ocorrências dos termos em cada país

Fonte: tradução pela autora de BEAK, J. et al, 2013. p.34.

Brasil	**Coreia do Sul**	**Espanha**	**Estados Unidos**
Representação do conhecimento	Classificação Decimal Coreana	*Web* semântica	Classificação
Gestão do conhecimento	Classificação Decimal de Dewey	Ontologias	Recuperação da informação
Ontologias	RDA	Tesauros	Ciência da Informação
Indexação automática	Classificação decimal *Nippon*	Organização do conhecimento	Sistemas
Sistemas de classificação	Regras de catalogação Coreana	Arquitetura da informação	Recuperação
Organização da Informação	FRBR	Gestão do conhecimento	Ciência
Ciência da Informação	OPAC	Recuperação da informação	*Web*
Instrumentos de gestão do conhecimento	Catálogos de bibliotecas	Sistemas de informação	Modelos
Catálogo *online*	*Metadata*	Internet	Informação
Semiótica	Cabeçalho de assunto	Sistemas de organização do conhecimento	Organização do conhecimento
Terminologia		Sistemas de informação de direito	Busca
		Metadata	Categorização
		SOC	Conhecimento
		Espanha	
		Bibliotecas universitárias	
		Web 2.0	

Um dos primeiros trabalhos realizados no Brasil sobre um panorama da área de Organização do Conhecimento foi publicado em 2008, pela professora Mariângela Fujita, com o título “Organização e representação do conhecimento no Brasil: análise e aspectos conceituais e da produção cientifica do ENANCIB no período de 2005 a 2007”. Neste trabalho, Fujita (2008) apresenta um panorama da pesquisa nesta área no Brasil, identificando as linhas e temas de pesquisa, seus fundamentos teóricos e metodológicos, e seu desenvolvimento cientifico. Para isso, faz um estudo sistemático sobre a produção do Grupo de Trabalho Organização e Representação do

Conhecimento (ORC) -GT2-, do Encontro Nacional da Associação Nacional da Pesquisa (ENANCIB), realizado anualmente pela Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB), produzidos entre os anos de 2005 e 2007. Em sua análise, a autora aponta a incipiência de estudos de articulação cientifica com as áreas interdisciplinares e ressalta o surgimento de temas tais como ontologia, taxonomias, *web* semântica e necessidade de aprofundar os estudos sobre os sistemas de classificação para torná-los mais flexível para atender às necessidades da área.

Também no ano de 2008, Guimarães apresentou o artigo intitulado “A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da *International Society for Knowledge Organization* (ISKO)”, em que faz uma análise da interlocução da organização do conhecimento, dentro do universo da ISKO, a partir das três correntes: catalogação de assunto, de influência norte americana; indexação, de influência inglesa; e análise documentária, de influência francesa. Dentro desse contexto, o autor faz uma análise das pesquisas e estudos apresentados no GT2 da ANCIB, e do capítulo brasileiro da ISKO, e conclui que os estudos realizados nas escolas brasileiras de BCI, nesta área, têm apontado para uma equidade na realização de estudos e pesquisas, sem perder de vista as abordagens internacionais. Como necessidade de desenvolvimento na área, Guimarães (2008) aponta: (1) necessidade de reflexão sobre os aspectos teóricos e metodológicos que contemplem a dimensão interdisciplinar sem perder seu foco principal de tratar a informação para se ter uma recuperação eficiente; (2) reflexão sobre a atuação profissional diante dessa necessidade, e (3) sugere que dentro do contexto da ISKO não se deve perder de vista como a área de Tratamento Temático da Informação (TTI) vem construindo seu referencial teórico-metodológico.

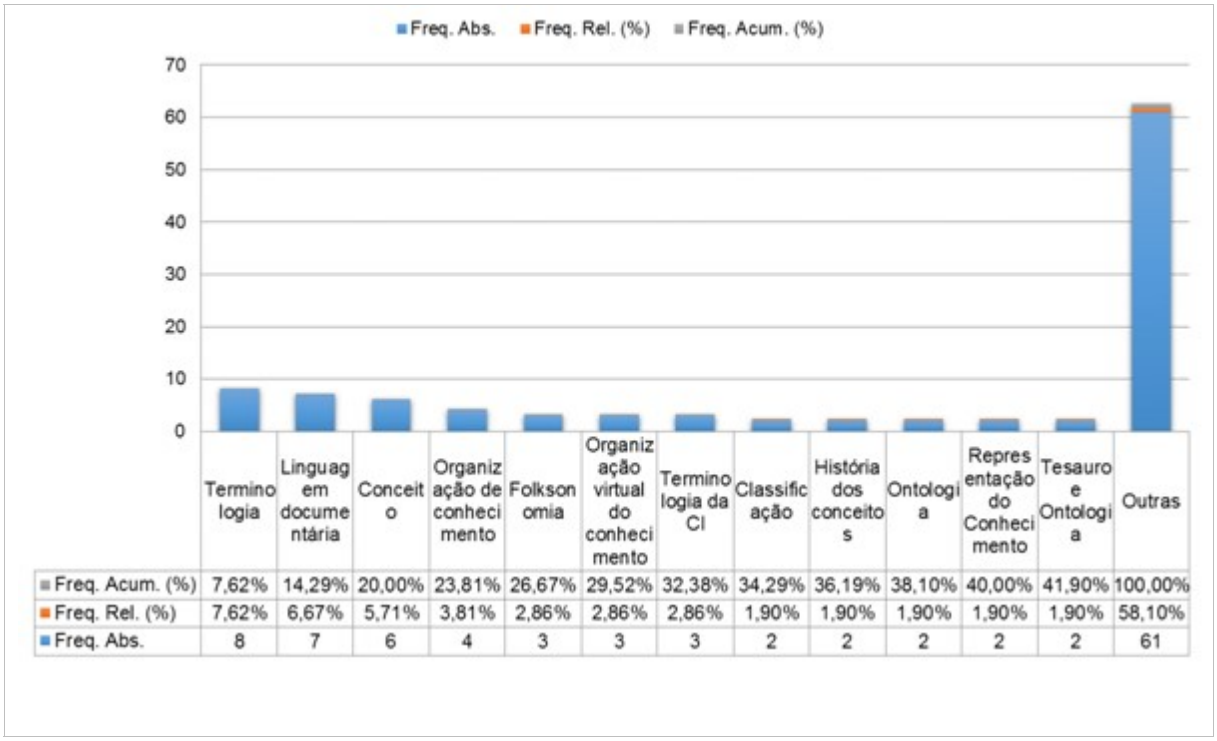
A Profª. Dra. Hagar Espanha Gomes publica em 2009 o artigo intitulado “Tendências da pesquisa em organização do conhecimento”, desenvolvido a partir da palestra apresentada no II Seminário de Pesquisa em Ciência da Informação, realizada pela Coordenação de Ensino e Pesquisa e Tecnologia de Informação do IBICT, no Rio de Janeiro, em 2007. A autora descreve um panorama geral da área e aponta dois aspectos importantes para pesquisa: (1) a abordagem teórico-filosófica voltada para questões epistemológicas para analisar um domínio com objetivo de elaboração de sistema e tabelas de classificação, e (2) os métodos empregados na elaboração destes sistemas. Além disso, Gomes (2009) sugere outros estudos

importantes de se apontar: (1) necessidade de se desenvolver uma base teórica sólida para as atividades da BCI, com possibilidade de uma teoria mais abrangente que possa subsidiar os estudos entre a teoria e a prática; (2) a multidisciplinaridade e o inter-relacionamento que ocorre entre os campos de pesquisa reforçam a necessidade de estudo das tradicionais tabelas de classificação e o desenvolvimento de outros sistemas, com maior hospitalidade e flexibilidade, para que permitam a representação de novos conceitos em suas estruturas vindo dessas áreas, (3) sugere a abordagem analítica do domínio, a teoria dos níveis integrativos, e o método de facetas como alternativas para esses processo, (4) pesquisas sobre a utilização das tabelas Classificação Decimal de Dewey (CDD) e a Classificação Decimal Universal (CDU) como instrumentos de apoio para representação e recuperação da informação na *web*.

Em 2014, Guimarães et al. publica o artigo “A dimensão temática da pesquisa em organização do conhecimento: uma análise de domínio dos congressos nacionais e regionais da ISKO”, no qual os autores fazem uma pesquisa exploratória utilizando a técnica de análise de domínio nos artigos publicados nos congressos dos capítulos da ISKO - Brasil, ISKO - Espanha-Portugal e ISKO - Canadá-Estados Unidos, nos anos de 2011 e 2013. Por meio da análise dos artigos, os autores apontaram os assuntos que foram predominantes nos três Capítulos: (1) epistemologia, (2) análise de domínio, (3) processos e linguagem de indexação (incluindo a classificação).

O último artigo encontrado na literatura brasileira, na pesquisa bibliográfica realizada neste trabalho, foi “Autoria e pesquisa em Organização do Conhecimento: análise da produção científica em Ciência da Informação”, publicado por Bufrem; Freitas; Nascimento (2014). Os autores fazem um estudo de citação e cocitação em um *corpus* de 105 artigos da área, publicados entre 2003 e 2012. São apresentadas nestes estudos tabelas que demostram quais os periódicos mais profícuos na área Organização do Conhecimento, os autores mais representativos e citados na produção cientifica dessa área, os enfoques de pesquisa encontrado no *corpus* dos artigos e a temáticas mais frequentes desse *corpus*. Este último item está apresentado na TAB. 1 a seguir.

Tabela 1 – Temáticas mais frequentes do *corpus* de artigos sobre Organização do Conhecimento no Brasil (2003 – 2012)



	Terminologia	Linguagem documentária	Conceito	Organização de conhecimento	Folksonomia	Organização virtual do conhecimento	Terminologia da CI	Classificação	História dos conceitos	Ontologia	Representação do Conhecimento	Tesauro e Ontologia	Outras
Freq. Acum. (%)	7,62%	14,29%	20,00%	23,81%	26,67%	29,52%	32,38%	34,29%	36,19%	38,10%	40,00%	41,90%	100,00%
Freq. Rel. (%)	7,62%	6,67%	5,71%	3,81%	2,86%	2,86%	2,86%	1,90%	1,90%	1,90%	1,90%	1,90%	58,10%
Freq. Abs.	8	7	6	4	3	3	3	2	2	2	2	2	61

Fonte: (BUFREM; FREITAS; NASCIMENTO, 2014, p.160), adaptado pela autora.

Pela breve revisão de literatura apresentada acima, pode-se afirmar que a teoria e a pesquisa na área de Organização do Conhecimento são processos cíclicos, pois a teoria precisa da pesquisa para desenvolver e ser aplicada, ao mesmo tempo a pesquisa precisa da teoria para embasar, direcionar e aperfeiçoar futuros trabalhos que resultam em avanços e desenvolvimento da área. Svenonius (1981) termina seu artigo com a seguinte frase "...*theory without practice is empty, but practice without theory is blind".*

2.1 Tendências atuais da pesquisa em organização do conhecimento no Brasil

A informática trouxe muitos benefícios para a área de Organização do Conhecimento, contribuindo para facilitar o armazenamento, processamento e recuperação da informação. As pesquisas na área de Organização do Conhecimento tiveram contribuições para o seu desenvolvimento, principalmente, pela interseção das áreas da Ciência da Computação, da Linguística, da Terminologia e da Ciência Cognitiva, que remonta aos últimos anos da década de 50. Este avanço impactou primeiramente nas atividades relacionadas ao processo de indexação, com o surgimento de sistemas automatizados para criação de índices e processamento automático do texto, com os aportes da linguística, no processo de catalogação, com

a criação do formato MARC, pela *Library of Congress* (LC), que passa a ser utilizado efetivamente em 1966, e que foi considerado pela *International Standard Organization* (ISO) o primeiro projeto de automação internacional a estabelecer normas de descrição bibliográfica em forma mecanicamente legível, possibilitando o intercâmbio de dados. Por sua vez, o processo de classificação pode vislumbrar a possibilidade de criação e compatibilização de vocabulários controlados com ajuda do computador. Todos esses procedimentos impactaram diretamente a performance da busca e da recuperação da informação nos sistemas de recuperação da informação (SRI)

Desde então, vários experimentos e estudos têm sido realizados, como foi exemplificado pelos trabalhos descritos acima. Nota-se que a mudança da forma tradicional de organizar o conhecimento sofreu impacto ainda maior com a consolidação da Internet na década de 1980.

Não é objetivo, no entanto, deste trabalho, apresentar todo o histórico desses acontecimentos. Além disso, nem todas as questões anteriores de pesquisas foram ainda desenvolvidas ou obtiveram respostas satisfatória. Por isso, a necessidade de voltar o olhar para essas questões ainda é pertinente. Passa-se, portanto, a apontar quais as necessidades atuais das pesquisas na área, principalmente no Brasil:

1) A teoria e os métodos em Organização do Conhecimento necessitam ser mais bem articulados e fundamentados para dar consistência e coerência à área;
2) A característica interdisciplinar da área requer um mapeamento terminológico, para que ela possa demarcar seu campo epistemológico e evitar ambigüidades geradas pelos uso de termos nas áreas de interseção;
3) Estudos sobre possíveis soluções para a interoperabilidade, visando ao compartilhamento de recursos informacionais, dando ênfase ao desenvolvimento de esquemas e padrões de metadados, para facilitar o acesso a informações nos diversos repositórios digitais publicados na *Web*;
4) Estudos sobre o potencial de aplicação da teoria da classificação facetada na organização e representação, assim como na recuperação da informação na *Web*;
5) Investigações sobre o uso dos sistemas de classificação bibliográficas (CDD e CDU) como ferramentas potenciais para navegação e busca por meio de suas estruturas hierárquicas como alternativa pra recuperação da informação;

6) Pesquisas que possibilitem a compatibilização entre diferentes tesauros e identificação de métodos que permitam que eles sejam utilizados, acessados e manipulados *online;*
7) Estudos de como modelar e representar a multidisciplinaridade dos domínios do conhecimento em sistemas de organização do conhecimento (SOCs) tendo em vista as necessidades dos usuários;
8) Aprofundamento das pesquisas sobre a contribuição da área da Terminologia na construção de SOC, para modelar o domínio especifico antes da sua representação, com auxílio de definições conceituais para estabelecer correlações entre os conceitos oriundos de diferentes áreas do conhecimento;
9) Investigação sobre as aplicações na recuperação da informação em diferentes contextos, para apoiar a pesquisa exploratória;
10) Pesquisa sobre a potencialidade do modelo de dados *Simple Knowledge Organization System* (SKOS) para representar a estrutura de diferentes tipos de (SOCs), para permitir a publicação dessas ferramentas com seus dados interligados através da proposta *Linked Open Data*;
11) Maior integração entre as comunidades de pesquisadores da Organização do Conhecimento e áreas de interseção;
12) Avaliação da busca realizada por palavra chave *versus* busca realizada com auxílio de um vocabulário controlado em ambiente *Web*;
13) Estudos sobre a implementação do RDA (*Resource Description and Access*), que permite a representação e descrição de recursos digitais;

Estudos sobre indexação automática, desenvolvendo técnicas mais eficientes de extração de termos, com a contribuição da linguística computacional e a utilização da inteligência artificial.

3 Considerações finais

Pode-se afirmar que os avanços que ocorreram desde a década de 1950 até os dias atuais foram relevantes e marcaram o desenvolvimento da área de Organização do Conhecimento, tanto no armazenamento e na representação quanto na recuperação da informação. O armazenamento, antes feito somente nos discos

rígidos dos computadores com grande capacidade de processamento, agora se dá de maneira diversa, que modo que há a possibilidade de se acessarem arquivos, dados e aplicativos em qualquer lugar e a qualquer hora, utilizando-se tanto um computador como dispositivos móveis, desde que haja conexão com a Internet, com o armazenamento nas chamadas nuvens. Com o crescimento e desenvolvimento da Internet cada vez mais acelerados, surgem novos interesses de pesquisa em tecnologias da *Web* Semântica e da *Web* 2.0, com uma participação cada vez maior dos usuários, que fornecem metadados e contribuem para a marcação social, como a folksonomia. As necessidades dos usuários deveriam se tornar o ponto central de estudos na área, já que a recuperação da informação é o principal foco de toda a área da Organização do Conhecimento. Além da dependência do desenvolvimento tecnológico, quanto maior for a integração entre os profissionais da área de Organização do Conhecimento com outros profissionais das áreas interdisciplinares, maior será o avanço nas pesquisas.

Referências

BEAK, J. et al. International comparative domains analysis in knowledge organization reseach topics in four countries - Brazil, South Korea, Spain and the United States. *North American Symposium of Knowlodge Organization,* v. 4, n.1, p.30-39. 2013. Disponível em: <https://journals.lib.washington.edu/index.php/nasko/issue/view/1012>. Acesso em: 2 ago. 2015.

BEGHTOL, C. Knowledge domain: multidisciplinar and bibliographic classification systems. *Knowledge Organization*, v.25, n.1/2, p.1-12. 1998.

BRÄSCHER, M.; CAFÉ, L. Organização da Informação ou Organização do Conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo, Anais. São Paulo: ANCIB, 2008. Disponível em: <http://www.enancib2008.com.br>. Acesso em: 31 jul. 2015.

BUFREM, L. S.; FREITAS J. L.; NASCIMENTO, B. S. Autoria e pesquisa em Organização do Conhecimento: análise da produção cientifica em Ciência da Informação. *Revista da Faculdade de Biblioteconomia e Comunicação da UFRGS*, v.20, n.3, p.150-165. Ed. Especial, 2014.

DAHLBERG, I. Current trends in knowledge organization. In: MARCO, F. J. García (Coord.). *Organización del conocimiento en sistemas de información y documentación 1*. Actas del I Encuentro de ISKO-España, p.7-25, 1993. Madrid: Librería General.

DAHLBERG, I. Knowledge organization: a new science? *Knowledge Organization*, n.33, v.1, p.11-19. 2006.

DAHLBERG, I. Knowledge organization: its scopes and possibilities. *Knowledge Organization*, Frankfurt, v.20, n. 4, p.211-222, 1993.

DIAS, E. W. Ensino e pesquisa em ciência da informação. *DataGramaZero - Revista de Ciência da Informação,* v.3, n.5, out./ 2002. Disponível em: <http://www.dgz.org.br/out02/F_I_art.htm>. Acesso em: 2 ago. 2015.

FUJITA, M. L. S. A. Organização e representação do conhecimento no Brasil: análise de aspectos conceituais e da produção cientifica do ENANCIB no período de 2005 a 2007. Tendências da Pesquisa Brasileira em Ciência da Informação, v. 1, n.1. 2008. Disponível em: <http://inseer.ibict.br/ancib/index.php/tpbci/article/view/4/13>. Acesso em: 27 jul. 2015.
GOMES, H. E. Tendências da pesquisa em organização do conhecimento. *Tendências da Pesquisa Brasileira em Ciência da Informação***,** v. 2, n. 1. p.60-88. 2009. Disponível em: <http://inseer.ibict.br/ancib/index.php/tpbci/issue/view/2/showToc>. Acesso em: 1 ago. 2015

GUIMARÃES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). *Revista Ibero-Americana de Ciência da Informação*, v.1, p.77-99, 2008. Disponível em: <http://periodicos.unb.br/index.php/RICI/article/view/2761>. Acesso em: 1 ago. 2015.

GUIMARÃES, J. A. C. et al. A dimensão temática da pesquisa em organização do conhecimento: uma análise de domínio dos congressos nacionais e regionais da ISKO Brasil, Espanha e América do Norte. *Scire*. v.20, n.2, p.19-25, jul.-dez. 2014. Disponível em: <http://ibersid.eu/ojs/index.php/scire/article/view/4177>. Acesso em: 1 ago. 2015.

HJORLAND B., ALBRECHSTEN H. An analysis of some trends in classification research. *Knowledge Organization,* v.26, n.3, p.131-139, 1999.
IBEKWE-SANJUAN, F.; SANJUAN, E. Knowledge organization in the last two decades: 1998-2008. In: GNOLI, C.; MAZZOCCHI (Ed.). Paradigms and conceptual systems in Knowledge Organization, *Advances in Knowledge Organization*, v.12, p.115-121, 2010.

LÓPEZ-HUERTAS, M. Some current research questions in the field of Knowledge organization, *Knowledge Organization*, n. 35, n.2-3, p.113-136, 2008.
MCILWAINE, I. C. Trends in knowledge organization. *Knowledge Organization*, n. 30, v.2, p.75-86. 2003.

McILWAINE, I. C., WILLIAMSON N.J. International trends in subject analysis. *Knowledge Organization*, v. 26, n.1, p.23-29,1999.

SAUMURE, K.; SHIRI, A. Knowledge organization trends in library and information studies: a preliminary comparison of the pre- and post-web eras. *Jornal of information science* v.34, n.5, p.651-66, 2008. Disponível em: <http://jis.sagepub.com/content/34/5.toc>. Acesso em: 27 jul. 2015.

SMIRAGLIA, R. About knowledge organization: an editorial. *Knowledge Organization*, v.32, n.4, p.139-140. 2005.

SMIRAGLIA, R. Whither knowledge organization? An editorial. *Knowledge Organization*, v.33, n.4, p.8-10, 2006.

SVENONIUS, E. Directions for research in indexing, classification, and cataloging. *Library Resources & Technical Services,* v. 25, n.1, p. 88- 103. January/March, 1991. Disponível em: <http://downloads.alcts.ala.org/lrts/lrtsv25no1.pdf>. Acesso em: 25 jul. 2015.

A ISKO Brasil e os grupos de pesquisa em organização do conhecimento

Evelyn Orrico
Universidade Federal do Estado do Rio de janeiro
evelynorrico@unirio.br

1 Introdução

O papel da Organização do Conhecimento é fundamental porque, por seu intermédio, organizamos nossa compreensão de mundo. Dito de outra maneira, pode-se perceber que, por intermédio da concepção da organização do conhecimento, é possível identificar como os diversos grupos sociais se representam ao longo do tempo.

A título de ilustração, podemos recuperar uma informação curiosa: o pesquisador Freire, em sua coluna semanal no Diário do Amazonas, no dia 2 de agosto, publica em sua coluna um texto sobre a incidência de suicídios entre comunidades indígenas. Nesse texto ele assume as palavras de outro pesquisador para afirmar que sequer existe tradução para esse termo na maioria das línguas indígenas, o que nos faz inferir que a prática suicida ou não existia, ou não era reconhecida como ação identificável nessas comunidades.

Afirmamos isso, porque se reconhece que a relação entre as coisas do mundo e o próprio mundo se estabelece via linguagem. Se não existe termo para designar uma determinada ação que um grupo social desempenha no mundo, é porque essa ação não assumiu papel social digno de nota para esse grupo e, por isso, sem a necessidade de ser designada por termo criado especificamente para tal.

Em contrapartida, as ações são igualmente designadas em virtude da compreensão que as diferentes sociedades em distintos momentos históricos fazem dela. A título também de ilustração, podemos citar a demanda feita à Editora Melhoramentos, que edita o Dicionário Michaelis, por intermédio do site Change.org com vistas à alteração da definição do verbete matrimônio, segundo publicado por

Senra (2015). Até então, as acepções desse verbete definiam-no como 1. União legítima entre homem e mulher.

A petição online criada por Eduardo Santarelo justifica-se pelas alterações sociais ocorridas nas últimas décadas e que permitem a pessoas em relação homoafetiva estabelecerem uniões reconhecidas legal e judicialmente.

Ora, o que esses dois exemplos nos apontam é que o campo da organização do conhecimento e da representação da informação é campo de conhecimento que demanda assídua e contextualizada reflexão, tendo em vista as constantes transformações por que passam e estabelecem as relações sociais.

Importa, então, saber qual o impacto que uma associação sobre organização do conhecimento tem provocado no fortalecimento do próprio campo.

2 Panorama da ISKO Brasil

O Capítulo brasileiro da ISKO Brasil, cuja criação emanou-se dos pesquisadores que participavam do GT2 da ANCIB, Organização e Representação do Conhecimento, vinha sendo organizado desde 2005 e, finalmente, em 2007, durante o VIII ENANCIB, foi finalmente instalado e teve aprovado o seu estatuto.

Desde sua criação até este ano de 2015, a curva de associados vem crescendo, como se pode ver pelo gráfico 1 a seguir, com dados a partir de 2000.

Gráfico 1 – Associados entre 2009-2015

Em relação à categoria de associados, são três as possibilidades: profissional ativo, aposentado e estudante. E nos últimos quatro anos apresentam a seguinte distribuição, evidenciada no gráfico 2.

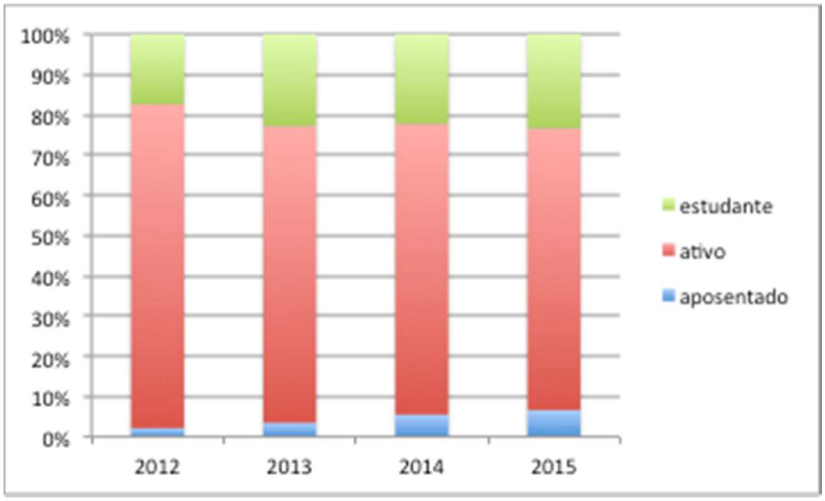


Gráfico 2 – Categoria de associados 2012-2015

A trajetória da Associação pode ser considerada como exitosa, tendo em vista que a curva de associados vem se mantendo ascendente. É igualmente digno de percepção que a balança entre aqueles que saem da vida profissional ativa devido à aposentadoria e os que entram na vida profissional, como os jovens estudantes, vem se mantendo em equilíbrio. O número de aposentados tem aumentado, mas o de estudantes também. Chamo atenção para esse dado porque a presença de jovens profissionais pode refletir um impacto positivo na continuidade da Associação.

Considerando que a maior parte de nossos associados é composta por profissionais que estão em atividade, procuramos refletir sobre o impacto que esse corpo de associados, e consequentemente da ISKO-Brasil, possa estar provocando na produção científica do próprio campo.

3 Perspectiva de impacto

O caminho adotado para perceber o impacto que a ISKO Brasil possa estar provocando na produção acadêmica do campo foi levantar os grupos de pesquisa do CNPq, realizando uma busca pelos termos organização e representação do conhecimento que se relacionasse com nossa área de interesse e com os nossos pesquisadores. Os resultados são de agosto de 2015, portanto recentes, e esse levantamento nos apresentou dados que nos permitem pensar em ações futuras.

Em primeiro lugar, fizemos a busca pela palavra-chave organização do conhecimento, sem qualquer espécie de filtro e sem exigir as palavras exatas. Tal

levantamento nos apresentou 101 grupos registrados no CNPq de várias áreas do conhecimento: Ciências da Saúde (2); Ciências Exatas e da Terra (32); Ciências Humanas (18); Ciências Sociais Aplicadas (38); Engenharias (3); Linguística, Letras e Artes (8). O que se depreende desse resultado é que esses termos instigam pesquisas em diversos âmbitos conceituais.

No entanto, ao estabelecermos a busca pelos termos organização do conhecimento, filtrando por palavras exatas, esse quantitativo cai para 25 grupos, com a seguinte distribuição por áreas: Ciências Exatas e da Terra (2); Ciências Sociais Aplicadas (23). Esse resultado ratifica um âmbito teórico-conceitual que insere as pesquisas sobre organização do conhecimento naquele rol de disciplinas que pertencem à grande área do conhecimento das Ciências Sociais Aplicadas, em que se inserem a Biblioteconomia, Arquivologia, Ciência da Informação.

Refletindo especialmente sobre a relação entre a ISKO Brasil e a criação de grupos de pesquisa dedicados ao tema de organização do conhecimento, refinamos os levantamentos, no intuito de identificar o período de tempo em que os grupos foram criados. O Diretório do CNPq nos permite as seguintes possibilidades de levantamento de criação dos grupos: há menos de 1 ano; entre 1 e 4 anos; entre 5 e 9 anos; entre 10 e 14 anos; e 15 anos ou mais.

Os resultados encontrados incluem grupos de pesquisa atualizados no último ano e grupos considerados não atualizados, ou seja, os que não sofreram atualização nos últimos 12 meses.

Encontramos os seguintes dados, dispostos na tabela 1:

Tabela 1 – Período de criação de grupos de pesquisa e seus componentes

Período de criação	Quantidade de grupos de pesquisa	Grupos com presença de associados	Líder associado	2º líder associado
Menos de 1 ano	0			
1 a 4 anos	10	5	4	1
5 a 9 anos	7		4	3
10 a 14 anos	4	2		

15 anos ou mais	4

Nos grupos de 1 a 4 anos, dos 10 grupos existentes, cinco apresentam sócios da ISKO, sendo quatro como líderes e um como 2º líder, em que 1º não é associado.

Nos grupos de 5 a 9 anos, dos 7 grupos criados, 3 apresentam líderes como sócios e 4 estão compostos com o 2º líder associado), tendo a seguinte distribuição: 2 grupos têm 1º e 2º líderes associados e 2, somente os 2º líderes o são.

Nos quatro grupos criados entre 10 e 14 anos, 2 têm associados como primeiro líder.

Dos quatro grupos criados há 15 anos ou mais, três apresentam líderes associados e um deles, além do líder, o 2º líder também faz parte da ISKO-Brasil.

Considerando os oito anos de criação da ISKO Brasil, os conjuntos de grupos que mais poderiam ter sido impactados seriam os três primeiros de nossa listagem: entre 5 e 9 anos; entre 1 e 4 anos e há menos de 1 ano.

Excetuando-se este último ano, em que não foi criado nenhum grupo nessa temática, os dois conjuntos seguintes, entre 1 e 4 anos e entre 5 e 9 anos, foram os de maior crescimento, indicando que profissionais de interesse aproximados agregaram esforços em prol de objeto comum. Por isso sugerimos tentar compreender esse dado pelo associativismo como possível mediador desse impacto.

4 Associativismo

Refletimos sobre o associativismo no intuito de compreender, ou eventualmente, propor políticas de fortalecimento para a ISKO Brasil, tendo em vista que a proposta pensada por Tocqueville (1998) pode contribuir para esse tipo de organização. O pressuposto dessa forma de organização é a participação de todos os participantes nas tomadas de decisão engendrando a forma criativa de buscar de forma conjunta o objeto dos seus interesses compartilhados.

Ele faz sua reflexões pautado na sociedade americana e seu movimento associativo, ressaltando que esse movimento resulta do agrupamento de pessoas sob um interesse comum a partir de uma sociabilidade formal e institucionalizada com algum poder político. Ele enfatiza que tal prática seja de colaboração e solidariedade.

Isso nos leva a pensar, como Fonseca (2008) que, ao estudar as associações com personalidade jurídica no Rio de Janeiro entre 1903 e 1916, destaca que seus

associados, reunidos em torno dos mesmos interesses, não só convergem esforços, mas também dão visibilidade à sua luta, configurando-se, então, interlocutores mais fortes no debate com o restante da sociedade e com o Estado, o que certamente amplia as possibilidades de mudança da realidade social. O objetivo central da ISKO Brasil é propiciar a interlocução entre seus associados, não especialmente estabelecer relação com o Estado, no entanto ela pode contribuir para representar as práticas sociais e com isso ajudar a sedimentar as transformações da realidade social.

A interlocução entre os associados evidencia-se nos três os objetivos do capítulo brasileiro da ISKO (http://isko-brasil.org.br) que, a nosso ver, coincidem com os pressupostos gerais do associativismo, que é fazer convergir interesses comuns e congregar esforços para a resolução de problemas:

- Promover a pesquisa, o desenvolvimento e aplicações de sistemas de organização conceituais do conhecimento que promovam o estudo dos aspectos filosóficos e semânticos da estrutura do conhecimento;
- Proporcionar os meios de comunicação e redes em organização do conhecimento para os seus associados; e,
- Funcionar como ponto de rede entre instituições nacionais e internacionais que trabalham com questões relacionadas à organização conceitual e à dinâmica do conhecimento.

Ainda segundo Fonseca (2008), as associações viam suas publicações como meios eficientes para atingir seus associados. Silva (2013) chega a afirmar que as essas publicações são fontes privilegiadas que retroalimentam a sua própria constituição.

Se no caso estudado por Silva (2013), o associativismo convergia para a institucionalização do campo arquivístico, no caso da ISKO Brasil, a criação da associação objetivou agregar pesquisadores especificamente em torno da temática ampla e multifacetada que é a Organização do Conhecimento.

5 Discussão e conclusão

O levantamento inicial dos grupos registrados no CNPq mostrou que em praticamente todas as áreas do conhecimento há interesses de pesquisa que perpassam os termos organização e conhecimento, mas que podem estar calcados

nos mais diversos contextos conceituais, como se pode ver pela inserção dos 101 grupos recuperados na busca: Ciências da Saúde (2); Ciências Exatas e da Terra (32); Ciências Humanas (18); Ciências Sociais Aplicadas (38); Engenharias (3); Linguística, Letras e Artes (8).

No entanto, quando a busca sofre a restrição do filtro “busca por palavras exatas”, esse número cai para 25 e as áreas de Conhecimento também ficam restritas a duas - Ciências Exatas e da Terra e Ciências Sociais Aplicadas - , sendo esta última a que concentra a maior parte dos grupos, apontando para a especificidade conceitual do sintagma organização do conhecimento.

Considerando as premissas do associativismo e refletindo sobre os dados encontrados, percebemos que a criação do capítulo brasileiro da ISKO estimulou e favoreceu a reunião de pesquisadores em prol do aprofundamento das questões do campo da Organização do Conhecimento. Isso pode ser visto pelo crescimento do número dos grupos de pesquisa registrados no CNPq nos anos próximos à data de criação.

No entanto, chamou-nos atenção a presença ainda discreta, embora em curva levemente ascendente, de estudantes na Associação. Sua presença seria um indicativo mais efetivo da força renovadora – e ao mesmo tempo fomentadora - de renovação do campo, garantindo-lhe uma perspectiva de maior vitalidade.

Outro dado digno de nota é a ausência de criação de grupos de pesquisa nesse campo nos últimos doze meses. Uma perspectiva possível é que os grupos criados há menos de 9 anos, período que compreende a criação da Associação, estejam agora começando a colher os frutos da maturidade dos novos membros formados no seio dos grupos mais antigos. Esperemos os próximos anos para constatar — ou não — essa possibilidade.

No entanto, como a busca realizada objetivava identificar os grupos formados, é necessário a partir de agora refinar as buscas, identificado, por exemplo, dentre os grupos aqueles que estão atualizados, produtivos, configurados.

Note-se igualmente que, dos grupos criados há mais tempo, embora em número pequeno - somente 4 -, três têm como líder, e um também o 2º líder, como nossos atuais associados. São grupos criados por pesquisadores que tradicionalmente já trabalhavam na área e continuam produzindo no campo da Organização do Conhecimento, alguns deles tendo participado da reunião que

manifestou “fortemente o interesse em criar o Capítulo brasileiro da ISKO”, conforme ata publicada no sítio eletrônico http://isko-brasil.org.br, na aba Memória.

Por fim, à semelhança do que afirmou Silva (2013) em relação à retroalimentação que publicações das Associações faziam em suas próprias convicções, talvez a introdução de uma política mais consistente de incentivo à publicação consiga fortalecer de maneira mais contundente o trabalho que vem sendo realizado até então na ISKO Brasil.

Sabemos todos que o trabalho solidário dos envolvidos em propostas de natureza associativa precisa estar em constante avaliação e promover frequentemente revisão das rotas adotadas. Essa mesa de perspectivas em Organização do Conhecimento deu-nos a oportunidade de refletir sobre o próprio caminhar associativo que a ISKO Brasil tem feito nesses seus oito anos de efetiva existência.

Referências

FONSECA, Vitor Manoel Marques da. **No gozo dos direitos civis**: associativismo no Rio de Janeiro, 1903-1916. Rio de Janeiro: Arquivo Nacional, Niterói: Muiraquitã, 2008.

FREIRE, José Ribamar Bessa. Como matar índio com giz e apagador. **Diário do Amazonas**, 2 de agosto de 2015. http://www.taquiprati.com.br/cronica.php?ident=1155ABNT 6023 e 10520. Acessado em 2/8/2015.

SENRA, Ricardo. Dicionário Michaelis muda definição de casamento após pressão online - Da BBC Brasil em Londres, 2 julho 2015. http://www.bbc.com/portuguese/noticias/2015/07/150702salasocialdicionariomichaeliscasamentors>. Acessado em 2 de agosto de 2015

SILVA, Eliezer Pires. **Memória e discurso do movimento associativo na institucionalização do campo arquivístico no Brasil (1971-1978**). 2013. Tese de Doutorado (Doutorado em Memória Social). Programa de Pós-Graduação em Memória Social, UNIRIO, Rio de Janeiro, 2013.

TOCQUEVILLE, Alexis de. **A democracia na América**. 4. ed. Belo Horizonte: Itatiaia, 1998.

Perspectivas da pesquisa em organização do conhecimento: reflexões a partir de trabalhos acadêmicos do PPGCI IBICT-UFRJ

Rosali Fernandez de Souza
Instituto Brasileiro de Informação em Ciência e Tecnologia
rosali@ibict.br

1 Introdução

Este texto foi escrito para apresentação na Mesa Redonda “Perspectivas da pesquisa em Organização do Conhecimento no Brasil” na reunião da III ISKO Brasil, realizada em Marilia SP, em 2015.

Buscando trazer uma reflexão em torno das principais questões do evento voltadas para a Organização do Conhecimento, quais sejam:

> Qual o papel da Organização do Conhecimento e seu compromisso social em relação aos desafios da diversidade cultural? Qual a participação do sujeito contemporâneo na multiplicidade de ações requeridas para a organização do conhecimento? Quais as perspectivas epistemológicas, tecnológicas e sociais que se colocam para o ensino, a pesquisa e as práticas em organização do conhecimento para fazer frente a um mundo permeado por uma cultura plural? Como vem se construindo o domínio da organização do conhecimento no universo científico? (http://isko-brasil.org.br/?page_id=21 acesso em 30 de junho 2015).

Considerei que seria uma oportunidade importante para eu mesma refletir como orientadora e pesquisadora e, ao mesmo tempo, dar a conhecer à comunidade reunida a produção acadêmico-científica do PPGCI IBICT-UFRJ em Organização do Conhecimento como contribuição de perspectivas de pesquisa nas questões propostas para discussão no evento.

1 Organização do Conhecimento e seu compromisso social em relação aos desafios da diversidade cultural

Na primeira questão: Qual o papel da Organização do Conhecimento e seu compromisso social em relação aos desafios da diversidade cultural? menciono o trabalho de dissertação de mestrado do aluno Rodrigo Piquet Saboia de Mello, por mim orientado, concluído em 2015. O título do trabalho: *Um olhar classificatório do acervo imagético das expedições científicas de Darcy Ribeiro aos índios Urubu-Kaapor no Museu do Índio na identificação de elementos da cultura indígena e da diversidade étnica-cultural brasileira* (PIQUET, 2015) – titulo um tanto extenso, mas explicita os elementos considerados chave do problema de pesquisa

Tendo a classificação como o foco que permeia a pesquisa, apresenta como corpus de análise o acervo imagético de fotos, arquivadas no Museu do índio (RJ), das expedições científicas aos índios Urubu-Kaapor que contaram com a participação do antropólogo Darcy Ribeiro - quando etnólogo da Seção de Estudos (SE) do Serviço de Proteção aos Índios (SPI), agência indigenista oficial do Estado Brasileiro.

Podemos considerar que, pelo menos cinco, são os fatores preponderantes a ressaltar na observação do corpus de análise do acervo das fotografias em questão: o olhar do antropólogo, o olhar do fotógrafo, o olhar do profissional da informação, o olhar do sistema de informação do Museu do Índio, onde o acervo está arquivado, e o olhar do usuário. Esses diversos olhares refletem necessariamente pontos de vista diferentes que influenciaram na idealização da foto para objetivos antropológicos e culturais e para os objetivos institucionais, na definição do enquadramento da foto enquanto arte de fotografar e na identificação e seleção dos elementos para as indexações temática e técnica. Ou seja, todos são aspectos a serem levados em consideração nos processos de representação e organização dos itens fotográficos das expedições para as necessidades de busca, acesso recuperação por possíveis usuários, assim como para a preservação do acervo no Museu do Índio.

As análises das fotos evidenciaram elementos relevantes da cultura indígena dos índios Urubu-Kaapor, tais como: Alimentação, Doença, Festa da Nominação e Arte Plumária. Quanto à diversidade étnica-cultural brasileira foram identificadas como exemplos de categorias temáticas: Oratório, Negros, Sertanejos e Festa do Bumba-meu-Boi.

Outra categoria temática identificada como relevante aos propósitos da indexação das fotos das expedições foi a de O Homem Cordial - conceito criado por Darcy Ribeiro. Na análise das fotos essa categoria foi representada pelas duas

personalidades chave que participaram das expedições: o pesquisador inglês, à época doutorando em Oxford, Francis Huxley quando da Segunda Expedição Científica e, o próprio Darcy Ribeiro, coordenador das duas expedições.

Quanto ao antropólogo social Francis Huxley é de se ressaltar o título de sua obra, produto da expedição, intitulada *Selvagens Amáveis* que revela duas características captadas dos índios Urubu Kaapor: serem selvagens e, ao mesmo tempo, amáveis. Além disso, é importante também ressaltar o interesse desse pesquisador em investigar o sistema de parentesco dos índios.

Nesse sentido, é importante relatar uma constatação comentada pelo aluno durante nossas discussões que bem ilustra, a meu ver, um significado de compromisso social da organização do conhecimento em relação a desafios da diversidade cultural que coincide com o interesse demonstrado por Francis Huxley. Relatou o aluno, funcionário do Museu do Índio, que em conversa com um índio membro da tribo Urubu-Kaapor, ao mostrar o acervo das fotos das expedições no catálogo, e ao exemplificar o processo de indexação, tomou aleatoriamente um determinado item fotográfico do acervo do Museu. Ao ver tal foto, o índio teria, com emoção, identificado o seu avô, cacique da tribo – um olhar não previsto pelo antropólogo, pelo fotógrafo e pelo profissional da informação - ou seja, a sutileza de um olhar do ponto de vista simplesmente humano, mas que reflete aspecto de organização social e cultural da tribo.

O antropólogo-etnólogo Darcy Ribeiro foi fotografado em vários momentos, e em séries fotográficas, em imagens que sugerem familiaridade e proximidade com os índios Urubu-Kaapor, retratando uma relação cordial e de amizade com os índios – objetivo que vai de encontro às próprias finalidades das expedições e que também indica a relevância dessa categoria temática.

A contribuição de Kossoy, citada no trabalho de Piquet, sobre análise iconográfica:

> A análise iconográfica, no caso da representação fotográfica situa-se a meio caminho da busca do significado do conteúdo; ver, descrever e constatar não é o suficiente. É este o momento de uma incursão em profundidade na cena representada, que só será possível se o fragmento visual foi compreendido em sua interioridade. Para tanto, é necessário, a par de conhecimentos sólidos acerca do momento histórico retratado, uma reflexão centrada no conteúdo, porém, num plano além daquele que é dado ver apenas pelo verismo iconográfico. É este o estágio mais profundo da investigação, cujos limites não são

> cristalinamente definidos. Não raro o pesquisador se surpreende refletindo neste plano pós-iconográfico, buscando os elos para a compreensão da vida que foi. (KOSSOY, 1989, p.65 apud PIQUET, 2015, p.90).

Ilustra as reflexões feitas, abrindo perspectivas de pesquisa em organização do conhecimento.

2 O sujeito contemporâneo na multiplicidade de ações para a organização do conhecimento

O trabalho que nos faz refletir sobre a participação do sujeito contemporâneo na multiplicidade de ações requeridas para a organização do conhecimento foi a pesquisa da tese de doutorado de Fernando Ewerton, orientada por Geraldo Nunes (UFRJ) e por mim co-orientada, concluída em 2013, intitulada *Classificação política e discurso jornalístico: como as agências de notícia transnacionais representam candidatos a presidente do Brasil nas eleições de 1989 e 2010* (FERNANDEZ JUNIOR, 2013).

O material estudado consistiu da análise de conteúdo de 305 textos jornalísticos distribuídos por quatro agências de notícia transnacionais durante as eleições presidenciais brasileiras de 1989 e de 2010. As agências analisadas foram: a francesa Agence France Presse (AFP), a americana Associated Press (AP), a espanhola EFE, e a britânica/canadense Reuters. Como objetivo, buscou identificar critérios de relevância aplicados por agências de notícia transnacionais na representação dos candidatos em dois momentos marcantes da política nacional.

Como comenta Fernandez Júnior:

> A representação da informação é preocupação comum a jornalistas e cientistas da informação, entre outras áreas e profissões. Ambos trabalham para fornecer textos/ documentos que satisfaçam o leitor/usuário em sua demanda por informação. E à medida que a compreensão dessa demanda fica mais complexa, com a proliferação de plataformas de distribuição digital, aumentam os desafios tanto para quem produz quanto para quem indexa e recupera informação, o que justifica uma aproximação maior entre os dois campos de estudo. (FERNANDEZ JUNIOR, 2013, p. 27).

O que indica perspectivas de pesquisa em organização do conhecimento na interrelação de questões que envolvem as ciências da Informação e da Comunicação.

Nesse trabalho, linguagem, significado e representação estão intrinsecamente relacionadas. Nas palavras de Orrico e Oliveira (2006, p.155), citadas por Fernandez Junior, “as coisas não significam em si mesmas, mas os enunciadores/usuários constroem o significado”, a partir de “práticas e processos simbólicos através dos quais a representação, o significado e a linguagem operam” e que,

> Nessa concepção, admite-se que os atores sociais usem o sistema conceitual de sua cultura para construir significado por algum sistema representacional, no caso a linguagem. É por esse sistema que os atores sociais fazem com que o mundo torne-se significável e é por ele que comunicam esse mundo a outros atores. (ORRICO; OLIVEIRA, 2006, p.155 apud FERNANDEZ JUNIOR, 2013, p.202).

Nesse sentido, o procedimento metodológico seguido, e que mais se aproxima do interesse da presente discussão, foi a criação da classificação dos tipos de termos utilizados nos textos jornalísticos na caraterização dos candidatos:

- Termo Político: associado ao candidato e/ou ao seu partido;
- Termo Partidário: que associa o candidato ao nome do partido;
- Termo Pessoal: associado à pessoa, não aplicável ao partido;
- Termo Funcional: associado a atividades, cargos e funções do candidato;
- Termo Relacional: em relação a outros candidatos e partidos.

A criação dessa classificação de natureza política para representar os candidatos a presidente do Brasil nas eleições de 1989 e 2010 foi fundamental, mesmo imprescindível, para a análise dos termos nos textos jornalísticos usados por agência e entre agências de notícias nas duas eleições, buscando encontrar semelhanças e particularidades na representação dos candidatos.

Nas palavras conclusivas da tese, o doutorado assinala que:

> [...] o olhar da Ciência da Informação oferece uma abordagem própria para o entendimento dos critérios de relevância que levam à predominância de determinados termos/conceitos em textos jornalísticos, que pode ser explorada em outros temas e discursos, neste e em outros campos de estudo, colaborando para a compreensão das representações que circulam na sociedade digital (FERNANDEZ JUNIOR, 2013, p. 205).

o que sugere perspectivas de pesquisa em organização e representação do conhecimento.

O prosseguimento do trabalho na mesma linha de pesquisa em estágio de pós-doutoramento objetivou ampliar e aprofundar a reflexão sobre os critérios de relevância aplicados na construção do discurso jornalístico no contexto internacional. O trabalho, intitulado *Relevância no Discurso Jornalístico: aspectos organizacionais, idiomáticos e culturais na representação dos candidatos a presidente do Brasil na eleição de 2010* (FERNANDEZ JUNIOR, 2015), foi por mim supervisionado e concluído em 2015.

A pesquisa de pós-doutorado investiga o assunto da relevância do discurso jornalístico, usando como metodologia parâmetros de relevância tradicionalmente usados em sistemas de recuperação da informação. Como objetivos do trabalho, Fernandez Junior assinala:

> […] avaliar se **os critérios de relevância das agências são estáveis enquanto outros aspectos** (idioma) **mudam**; se, como as agências, **veículos diferentes usam critérios similares para atribuir relevância**, aplicando **diferentes pesos** e **múltiplos critérios, usados interativamente**; se **diferentes classificações políticas, em idiomas diferentes, usam termos similares, mas aplicam diferentes pesos**; e se **há consistência de critérios entre agências e veículos de origem nacional, idiomática ou cultural comum.** (FERNANDEZ JUNIOR, 2015, p. 6, grifo do autor).

Como contribuição do trabalho, Fernandez Junior ressalta que

> [...] ao focalizar os critérios de relevância e as classificações políticas presentes no discurso jornalístico em sites de notícias internacionais, procuramos contribuir para o entendimento das representações encontradas em textos armazenados em bases de dados digitais, numa época em que a internet desempenha papel central na distribuição de informação em escala global (FERNANDEZ JUNIOR, 2015, p.54).

abrindo assim novas perspectivas para pesquisas em organização do conhecimento no contexto da sociedade digital.

3 Perspectivas epistemológicas, tecnológicas e sociais para o ensino, a pesquisa e as práticas em organização do conhecimento frente a um mundo permeado por uma cultura plural

Na caraterização dessa seção incluo a dissertação de mestrado de Thulio Pereira Dias Gomes, por mim orientada com co-orientação com Sgarbi Goulart (UERJ), defendida em 2015, intitulada*: A charge é o assunto: análise documentária de charge.* (GOMES, 2015).

Nessa investigação ressalto dois pontos como importantes para a presente discussão: considerar e analisar a Charge como um documento e investigar sobre possíveis correlações entre procedimentos de tratamento da informação em base de Análise Documentária aplicada à Charge.

Devido a imprecisões nas definições de desenhos de humor, a investigação tratou dos conceitos de Charge, Cartum, e Caricatura para identificar peculiaridades e especificidades de ordem cultural, observadas segundo aspectos etimológicos e de definições encontradas em dicionários de línguas - portuguesa, francesa e inglesa e em outras fontes.

Foi definido como Charge um “gênero discursivo de uso híbrido das linguagens verbal e imagética, caracterizado pela temporalidade marcada pela sátira e pela crítica referentes a determinado evento, em geral de natureza política” (GOMES, 2013, p. 26-27).

Na investigação da Charge na dimensão pragmática da informação foram analisadas as variáveis do processo global da análise documentária propostas por Lara (LARA, 1993), adaptadas ao contexto de análise documentária de charge. Como variáveis do processo global da análise documentária de Charge foram identificadas: a estrutura descritiva da charge, as informações brutas presentes na charge, os domínios do conhecimento interessados na charge, as instituições de custódia de charge, os usuários da charge e o estoque de conhecimento anterior do analista e sua formação ideológica (GOMES, 2015 p.134).

Como aspecto conclusivo GOMES destaca:

> A existência de variáveis da análise documentária é uma evidência de que esse processo de organização do conhecimento é histórico, ou seja, circunscrito em condições sociais específicas. Por consequência, a representação documentária, seja qual for o documento, não pode ser assumida como neutra e imparcial, o que nos permite inferir sobre a inviabilidade de uma proposta universal e atemporal para análise documentária de charge (GOMES 2015, p.134).

o que não só evidencia aspectos imprescindíveis a serem considerados quando do tratamento da análise documentária da charge, como amplia as perspectivas de

pesquisa em representação da informação e organização do conhecimento em diferentes tipos de documentos.

4 O domínio da organização do conhecimento no universo científico

Considerando o domínio da organização do conhecimento no universo científico, três trabalhos podem ser mencionados na investigação de áreas e campos de conhecimento das grandes áreas das ciências exatas, das ciências agrárias e das ciências humanas.

Na grande área das ciências exatas, menciono a tese de doutorado de Luana Farias Sales Marques intitulada *Integração Semântica de Publicações Científicas e Dados de Pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares*, por mim orientada tendo como co-orientador Luís Fenando Sayão (CNEN), defendida em 2014.

O conceito de Publicada Ampliada, correlacionado Publicações científicas e Dados de Pesquisa necessita de uma Integração Semântica em base de uma consistente de Organização e Representação da Informação. No mundo da e/Science essa efetivação é possível através de recursos tecnológicos e do desenvolvimento de padrões da Web Semântica, em base de relações conceituais consistentes.

A área de Ciências Nucleares é uma área de amplitude multi e interdisciplinar de abrangência científica tecnológica, política e social e que tem uma geração intensiva de dados científicos de todos os tipos, de natureza experimental e gerada por simulação.

Como aspecto conclusivo MARQUES ressalta que:

> [...] [n]ão obstante a forte presença de outras disciplinas no estudo de novos conceitos de publicações científicas para o ambiente da *eScience* - que tem a tecnologia como um fator determinante, principalmente as provenientes da Ciência da Computação - as teorias, metodologias e práticas postuladas pela Ciência da Informação e pela Biblioteconomia prevalecem como um componente ordenador e estruturante, e estão fortemente subjacentes a todos os estudos e aplicações da área, especialmente quando envolvem questões de representação e significado (MARQUES, 2014, p. 234).

que são questões típicas de organização do conhecimento

Para a presente discussão também é importante salientar as palavras de SALES:

> [...] a viabilidade de incorporação dos princípios e das teorias da Biblioteconomia e da Ciência da Informação para a organização do conhecimento técnico-cientifico no mundo da *eScience* (MARQUES, 2014).

o que indica a possibilidade de expansão da organização do conhecimento no papel de integração semântica no desenvolvimento de taxonomias e ontologias no âmbito da *eScience* - uma realidade que já estamos vivenciando.

A grande área das Ciências Agrárias é representada na presente discussão pela dissertação intitulada *Representação do Domínio da Agricultura no Contexto da Organização do Conhecimento* de Leandra Pereira de Oliveira, por mim orientada e defendida em 2015. O foco central da pesquisa foi direcionado aos sistemas de organização do conhecimento (SOCs) como fonte para conceituação e delimitação de um domínio de conhecimento.

O resumo da dissertação de OLIVEIRA apresenta a natureza da pesquisa, os objetivos, a metodologia seguida, os resultados encontrados e a contribuição do trabalho para o domínio da organização do conhecimento no mundo científico:

> Pesquisa exploratória que tem como objetivo principal investigar modelos representacionais, visando evidenciar particularidades da Agricultura como domínio do conhecimento. Para isso, foram realizadas pesquisas bibliográfica e documental para atender os objetivos específicos que são: identificar a origem etimológica e os aspectos históricos da Agricultura, investigar a área de Agricultura em classificações nacionais e internacionais de conhecimento e observar num contexto diferente de aplicação, outro modelo referencial da Agricultura. Apresentam-se como modelos representacionais do domínio da Agricultura: o *AGRIS/CARIS: Categorization scheme* da FAO, a Tabela de Áreas do Conhecimento do CNPq e a representação presente nos dados da pós-graduação brasileira na Grande Área 'Ciências Agrárias' da CAPES. A partir da descrição e análise dos modelos, observou-se proximamente a complexidade que cerca a representação de um domínio do conhecimento. No caso da Agricultura, a dificuldade reside na pluralidade de objetos de estudo que transcendem a disciplinaridade e a interdisciplinaridade do domínio. Essa característica, se por um lado, motiva e enriquece o desenvolvimento científico e tecnológico, por outro, leva a dificuldades de representação da Agricultura nos modelos referenciais e representacionais de organização do conhecimento. Conclui-se que a análise de diferentes modelos representacionais se coloca como uma abordagem para perceber as diferentes visões existentes no domínio por refletir a organização e representação do conhecimento praticada pela comunidade discursiva (OLIVEIRA, 2015).

As conclusões do trabalho evidenciam a importância da pesquisa em organização do conhecimento estar sempre atenta aos diferentes olhares do conhecimento científico pelas comunidades discursivas.

A grande área das Ciências Humanas é representada pela tese de doutorado: *O processo de comunicação científica no campo da defesa no Brasil: da geração do conhecimento à disponibilização da Informação* de Jaqueline Santos Barradas, orientada por Lena Vania Ribeiro Pinheiro (IBICT), defendida em 2015, que participei como membro da Banca Examinadora.

Esse trabalho, embora com o foco no processo de comunicação científica da geração à disponibilidade da informação, nos interessa para a presente discussão uma vez que evidenciou não só a dificuldade e a complexidade da definição e nomeação da Defesa como campo de conhecimento, como também as consequências das decisões tomadas nesse sentido para as análises realizadas e consequente interpretação dos resultados e das conclusões obtidas.

Uma importante constatação da pesquisa para a presente discussão nas palavras de Barradas, autora da tese:

> Outra constatação desta pesquisa foi verificar que não há consenso sobre o conceito da Defesa, nem sobre em qual área (do conhecimento) deve estar hierarquicamente vinculada. Há ainda muito a se delimitar, ressaltar, construir: suas origens e fronteiras no Brasil e no exterior: campo de conhecimento em que se insere; cruzamento epistêmico e interdisciplinaridade com outras áreas – como estudos estratégicos, ciências militares, CPRI (Ciência Política e Relações Internacionais) […] (BARRADAS, 2015, p. 183).

indica o papel delicado nas decisões de delimitação do escopo das áreas, campos e domínios do conhecimento retratados nas pesquisas em organização do conhecimento.

5 Considerações finais

Dos trabalhos acadêmicos relatados do PPGCI IBICT-UFRJ, alguns pontos emergiram como perspectivas de pesquisa em organização do conhecimento no Brasil.

A diversidade étnica-cultural brasileira e a respectiva documentação de atividades de natureza histórica, científica e social produzida em forma de documentos

de acervos de bibliotecas, museus e arquivos apresentam potencial rico e um desafio a ser explorado na pesquisa em organização do conhecimento. O caso específico relatado do acervo das fotos de expedições científicas, fruto da atuação do Museu do Índio, é um exemplo representativo de investigação dos olhares fotográfico, antropológico, sociológico e do profissional da informação que trazem implicações nas decisões das atividades de organização e representação do conhecimento.

A análise de textos jornalísticos abre perspectivas de pesquisa na interface da Ciência da Informação com a Ciência da Comunicação. O desenvolvimento de classificações políticas propiciou explorar a validade de critérios de relevância para o entendimento de fatores que influenciam a representação de informação em discursos jornalísticos , o que pode ser explorado em outros contextos de aplicação ampliando questões de pesquisa em organização do conhecimento.

A exploração da imagem como um documento e, como tal, objeto de análise documentária, amplia o foco da pesquisa em organização do conhecimento. No caso do desenho de humor, da "charge como o assunto", a aplicação de metodologias conhecidas não só enriqueceu, como suscitou o desenvolvimento de novos métodos para análise de narrativas de imagens, o que poderá ser ampliado para outros tipos de materiais, tendo a teoria e a prática da organização e a representação do conhecimento como ponto central de apoio.

O paradigma da *eScience*, considerado o 4º paradigma científico, é um desafio instigante para a organização do conhecimento. O exemplo de pesquisa acadêmica relatado revelou diversidades de aplicações da *eScience* voltadas para integração, recuperação e compartilhamento de informações de naturezas distintas. Destacou também a importância de *hiperlink*s, enquanto artefato tecnológico da Web semântica e da construção de taxonomias e ontologias que são típicos objetos de estudos e pesquisas em organização do conhecimento.

Os trabalhos que tratam de uma determinada área, campo ou domínio do conhecimento trazem desafiantes perspectivas para a pesquisa em organização do conhecimento. Os títulos referem-se a: Área das Ciências Nucleares, Campo da Defesa e Domínio da Agricultura. Um ponto a ser explorado é o próprio conceito de "área", "campo" e "domínio" de conhecimento. Nos três trabalhos foi mencionada a natureza multi e interdisciplinar do conhecimento o que dificulta a definição de uma área, de um campo e de um domínio para diferentes propósitos de pesquisa em organização do conhecimento.

Em todos os trabalhos ficou evidenciado o relevante papel da classificação como processo fundamental da natureza humana e como atividade insubstituível nos processos de disseminação, busca e recuperação da informação. Nesse sentido, torna-se inesgotável a identificação de questões de pesquisa em organização do conhecimento abrangendo diferentes tipos de documentos e meios de comunicação dos mais tradicionais aos mais sofisticados e já presentes na era digital.

Referências

BARRADAS, Jaqueline Santos. **O processo de comunicação científica no campo da defesa do Brasil: da geração do conhecimento à disponibilização da informação.** Tese de Doutorado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Lena Vania Ribeiro Pinheiro (IBICT). Co-orientador: Jorge Calvario dos Santos (UFF/Escola Superior de Guerra). Rio de Janeiro, 2015.

FERNANDEZ JUNIOR, Fernando Ewerton. **Classificação política e discurso jornalístico: como agências de notícia transnacionais representam candidatos a presidente do Brasil nas eleições de 1989 e 2010.** Tese de Doutorado em Ciência da Informação IBICT/UFRJ/FACC. Orientador: Geraldo Nunes (UFRJ). Co-orientador: Rosali Fernandez de Souza (IBICT). Rio de Janeiro, 2013.

FERNANDEZ JUNIOR, Fernando Ewerton. **Relevância no Discurso Jornalístico: aspectos organizacionais, idiomáticos e culturais na representação da informação dos candidatos a presidente do Brasil na eleição de 2010.** Pós-doutorado em Ciência da Informação IBICT/UFRJ/ECO. Supervisora: Rosali Fernandez de Souza (IBICT). Rio de Janeiro, 2015.

GOMES, Thulio Pereira Dias. **A charge é o assunto: análise documentária de Charge.** Dissertação de Mestrado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Rosali Fernandez de Souza (IBICT). Co-orientador: Paulo Sergio Sgarbi Goulart (UERJ). Rio de Janeiro, 2015.

GOMES, Thulio Pereira Dias. **Temas e questões em análise documentária de charge**. 2013. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia). Faculdade de Administração e Ciências Contábeis, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2013.
KOSSOY, Boris. **Fotografia & História**. São Paulo, Ateliê Editorial, 2012, p. 65

LARA, Marilda Lopes Ginez de. **A representação documentária: em jogo a significação**. 1993. 133 f. Dissertação de Mestrado em Ciências da Comunicação. Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 1993.

OLIVEIRA, Leandra Pereira de. **Representação do domínio da agricultura no contexto da organização do conhecimento.** Dissertação de Mestrado em Ciência

da Informação IBICT/UFRJ/ECO. Orientadora: Rosali Fernandez de Souza (IBICT). Rio de Janeiro, 2014.

MARQUES, Luana Farias Sales. **Integração semântica de publicações científicas e dados de pesquisa: proposta de modelo de publicação ampliada para a área de ciências nucleares.** Tese de Doutorado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Rosali Fernandez de Souza (IBICT). Co-orientador: Luís Fernando Sayão (CNEN). Rio de Janeiro, 2014.

MELLO, Rodrigo Piquet Saboia de. **Um olhar classificatório do acervo imagético das expedições científicas de Darcy Ribeiro aos índios Urubu-Kaapor no Museu do Índio na identificação de elementos da cultura indígena e da diversidade étnica-cultural brasileira.** Dissertação de Mestrado em Ciência da Informação IBICT/UFRJ/ECO. Orientadora: Rosali Fernandez de Souza (IBICT). Rio de Janeiro 2015.

ORRICO, E.G.D.; OLIVEIRA, C.I. de. A linguagem na construção do sentido: discurso e organização do conhecimento. In. GONZALEZ DE GÓMEZ, M.N.; ORRICO, E.G.D. (Orgs.). **Políticas de memória e informação: reflexos na organização do conhecimento**. Natal: Ed. UFRN, 2006, p.155

Perspectivas da pesquisa sobre organização do conhecimento no Brasil

Leilah Santiago Bufrem
Universidade Estadual Paulista
santiagobufrem@gmail.com

1 Introdução e evidências de interesse

A produção da literatura científica, elemento concreto de práticas institucionalizadas de pesquisa, constitui-se em processo dinâmico, gerador e renovador de domínios do conhecimento, mas principalmente desponta como objeto de investigação legítimo para que se compreenda o próprio campo de atuação científica. A possibilidade de reconhecer o que se produz em determinado campo, assim como seus elementos e marcas de distinção, tem motivado análises com pretensões de revelar a complexidade de práticas socialmente construídas e descobrir progressivamente as estruturas e modalidades de investigação científica, favorecendo aos pesquisadores olhares direcionados à compreensão do campo de produção intelectual em que atuam.

Ao ensejo do III Congresso Brasileiro em Organização e Representação do Conhecimento, ISKO-Brasil (*Conferenc*e 2015), a motivação deste estudo partiu da curiosidade sobre como se configura o domínio da Organização do Conhecimento (OC) na literatura periódica de Ciência da Informação (CI), mais especificamente em relação às suas dimensões (a) epistemológica, (b) aplicada e (c) social e política, conforme ementa sugerida para captar a abrangência dos interesses do evento.

A intenção de analisar como é representada essa temática na literatura periódica de CI, presente na Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação (Brapci), surgiu da curiosidade sobre as tendências perceptíveis na literatura pertinente ao tema do evento, nos últimos dez anos, no Brasil. Acredita-se que o autoconhecimento sobre o que é produzido no âmbito do domínio e sobre as transformações na pesquisa da área, cujos marcos permeiam o

processo científico, tem amparo significativo na literatura periódica. O desafio, entretanto, além da atualização sobre as comunicações científicas mais recentes, tem a ver com a produção expressiva no domínio da OC, cujas relações conceituais com outros domínios constituem-se em pressuposto deste trabalho. Assim, a intenção parte também do reconhecimento de que os pesquisadores do CNPq, especialmente do primeiro estrato, são representantes da excelência acadêmica, elemento determinante para a escolha de seus artigos como constituintes do corpus específico desta pesquisa. A motivação, ao incluir como sujeitos deste estudo os pesquisadores da OC, elaborando um corpus representativo de sua produção, origina-se na convicção de que essa análise pode revelar marcas de distinção na produção científica do domínio e da constatação de sua corresponsabilidade na produção científica coletiva.

A questão sobre como se configura o domínio de OC na literatura periódica de CI, representada na Brapci pelos pesquisadores do CNPq, passa a orientar este estudo, pelo qual se busca interpretar as relações evidenciadas nesse corpus, procurando situar os estudos no espaço constituído pela articulação dinâmica das dimensões sugeridas pelo evento. A concepção articuladora dessas dimensões, portanto, radica na complementaridade entre elas, eliminando-se a ideia de que sejam consideradas totalmente excludentes.

Justifica-se a escolha da Brapci como fonte de recuperação dos artigos, pelo reconhecimento de que esse instrumento permite a visualização da produção nacional de artigos de periódicos da área de Ciência da Informação. Estes, por sua vez, enquanto canais privilegiados de produção de conhecimento e modelos exemplares de uma prática relevante para o desenvolvimento da ciência, veiculam resultados de pesquisa e reflexão certificados por uma comissão científica, cujas atribuições consultivas e de assessoramento os respaldam.

2 Reconhecendo um domínio

As práticas científico-comunicativas de uma sociedade são afetadas pelas suas condições de existência e sua legitimação decorre das relações que os produtores de bens simbólicos mantém entre si (BOURDIEU, P., PASSERON, J.-C., 2014, p. 59). A possibilidade de que essas práticas ocorram entre as diferentes posições de um campo de produção cultural específico gera a diversidade nos modos de atuação em

busca da legitimidade no referido campo. Para compreendê-las sugere-se uma dupla leitura: a do campo propriamente dito, enfocando-se as práticas por ele legitimadas e a do campo político, que analisa as correntes teóricas nele dominantes. Aqui o campo se configura, conforme concepção de Bourdieu (1976), como um espaço simbólico, no qual as lutas dos agentes determinam, validam e legitimam representações, ou seja, um espaço objetivo de compromissos científicos engajados.

Mas a noção de campo, valiosa para o entendimento das posições e movimentos relacionais entre os pesquisadores, instituições e configurações temáticas, não é utilizada somente como recurso para o aperfeiçoamento da crítica ao fazer científico e sua avaliação. Isso porque os resultados da análise do campo científico facilitam a compreensão teórica dos objetos de investigação e da noção de domínio, necessária, segundo argumenta Lloyd (1995, p. 25), para que se delineie o processo de formação decorrente da produção de pesquisa científica. Esse tipo de análise facilita a determinação da coerência teórica e da adequação de procedimentos aos objetos de conhecimento que se constroem na medida em que os estudiosos contribuem para a sua inserção em determinado campo científico.

Portanto, a atual reflexão sobre as relações entre a institucionalização temática de um evento de OC e a produção científica concreta que a ele se integra amplia as possibilidades de compreensão desse domínio específico. Dai que a intenção de analisar a temática selecionada para o evento surgiu da curiosidade sobre suas tendências na literatura periódica brasileira. Embora reconheça que sempre haverá deficiências na explicitação dos saberes devido às novas informações e ao processo de mudanças sociais, Lloyd (1995, p. 18) argumenta que as metodologias e os conceitos gerais podem mostrar-se suscetíveis de permanente aperfeiçoamento em decorrência do pensamento analítico, incluindo o exame dos sistemas conceituais, da lógica de investigação e de raciocínio empregada por certas ciências e do modo como algumas delas avançam mais do que outras.

Como fundamento conceitual deste estudo, cuja intenção é analisar o domínio da OC na literatura, destaca-se a posição de Esteban Navarro e Garcia Marco (1995, p. 149, tradução nossa), considerada pertinente para as ponderações aqui apresentadas:

> [...] disciplina dedicada ao estudo e desenvolvimento dos fundamentos e técnicas do planejamento, construção, gestão, uso e avaliação de

> sistemas de descrição, catalogação, ordenação, classificação, armazenamento, comunicação e recuperação dos documentos criados pelo homem para testemunhar, conservar e transmitir seu saber e seus atos, a partir de seu conteúdo, com o fim de garantir sua conversão em informação capaz de gerar novo conhecimento.

A abrangência dessa definição, envolvendo os aspectos e as relações componentes da OC, aproxima-se da posição de Hjørland (2008, p. 86), referindo-se ao domínio de OC, como o estudo sobre a natureza e qualidade dos processos e sistemas de OC, relacionado com atividades como descrição de documentos, indexação e classificação em bibliotecas, bases de dados bibliográficas, arquivos, assim como de outros tipos de memória institucional.

Dai a sugestão de relacioná-la com as categorias representativas desse domínio, conforme o previsto na ementa do Congresso Isko 2015. Em consonância com essa concepção, Barité (2001, p. 38) destaca os subsídios teóricos que a OC proporciona aos temas “relacionados com o tratamento da informação e em particular o tratamento temático da informação, e de uma maneira menos específica - mas não menos importante - de gestão do uso social da informação”.

Ressalta-se, ainda, que por se tratar de uma análise da literatura da área, este estudo teve como suporte a pesquisa de Guimarães (2008) sobre a construção do referencial teórico-metodológico da OC, a partir de três correntes: a da catalogação de assunto (*subject cataloguing*), de influência norte-americana, a da indexação (*indexing*), de influência inglesa e a da análise documentária (*analyse documentaire*), de influência francesa.

3 Trajetória metodológica

A **primeira etapa** da pesquisa consistiu na identificação dos Pesquisadores Produtividade em Pesquisa Sênior e 1 (PQ1) dos níveis A, B, C e D do CNPq da área de Ciência da Informação. Foram identificados 21 pesquisadores (Gráfico 1), a partir da busca no site do CNPq, utilizando-se filtros para refino, tais como Grande Área (Ciências Sociais Aplicadas), Área (Ciência da Informação), Tipos de bolsa (Produtividade em Pesquisa) e Nível de Bolsa (Sênior e PQ1).

Gráfico 1 – Pesquisadores categoria 1 e Sênior, de CI – CNPq
Fonte: Site do CNPq (2015)

A contextualização e o reconhecimento do universo e dos sujeitos da pesquisa foram facilitados pela localização das instituições em que atuam, entre as quais se destaca o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), conforme observável no Gráfico 2, que também evidencia a maior concentração de bolsistas na região Sudeste. Este resultado relativo ao predomínio da região coincide com outros estudos nas diversas áreas do conhecimento (MENDES et al, 2006; BORTOLOZZI; GREMSKI, 2011) e tem sido atribuído à maior concentração de cursos de graduação e pós-graduação nessa região, que em 2008 era privilegiada com 61,1% dos programas com mestrado e doutorado entre todas as áreas (SANTOS; AZEVEDO, 2009). A peculiaridade deste estudo, entretanto, radica na presença de dois pesquisadores da região Nordeste no quadro representativo dos PQ1, o que proporcionalmente à quantidade de outras regiões, é significativo.

Ao interesse por esse tipo de relação correspondem hoje os estudos sobre redes sociais que, embora não tenham sido utilizados intensamente neste estudo exploratório, poderão contribuir para o delineamento das aproximações ou distanciamentos entre atores que, nesse âmbito podem ser pesquisadores, instituições, temas, períodos, ou outras realidades representadas nas redes. Sugerida também como ferramenta para o estudo dessas relações evidenciadas na produção científica, a ‘Análise de Redes Sociais’ (ARS) tem sido apresentada por Wasserman e Faust (1994) como forma de responder a questões de investigação nas áreas relacionadas com as ciências sociais, como recurso analítico adequado para as interpretações sobre posições, princípios, posturas, afinidades temáticas e metodológicas, abrangendo elementos que compõem a conjuntura científica de domínios específicos.

Gráfico 2 – Pesquisadores PQ1 e Sênior da Área de CI/Instituição
Fonte: Site do CNPq (2015)

A caracterização dos pesquisadores demandou a utilização de algumas ferramentas para a extração, compilação, visualização e organização dos dados:

Ferramenta	Propósito
Script Lattes	Baixar automaticamente os currículos Lattes do grupo de pesquisadores de interesse, compilando as listas de produções, tratando apropriadamente as produções duplicadas e similares, e representando as redes de colaboração científica entre os pesquisadores. Os dados foram obtidos na Plataforma Lattes do CNPq. (MENA-CHALCO; CESAR JUNIOR, 2009)
Web Qualis	Mapear os indicadores de qualidade dos artigos publicados em periódicos científicos por área do conhecimento (Ciências Sociais Aplicadas I), cujos critérios são estabelecidos pelos grupos especializados da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).
Vantage Point	Processar, classificar e representar a produção científica. Sua função foi a criação de matrizes de correlação que auxiliaram na representação das redes sociais científicas.
Microsoft Excel	Criar os gráficos e as tabelas da produção científica analisada.

Quadro 1 – Ferramentas em Relação aos Propósitos da Pesquisa
Fonte: Elaboração Própria (2015)

Especificamente na Plataforma Lattes, foram coletados, no dia 6 de julho de 2015, os dados de publicação dos pesquisadores PQ1 e Sênior no período de 2006 a 2015. A partir desses dados, foi possível chegar à rede de colaboração apresentada no gráfico 3, com todos os pesquisadores PQ1 e Sênior.

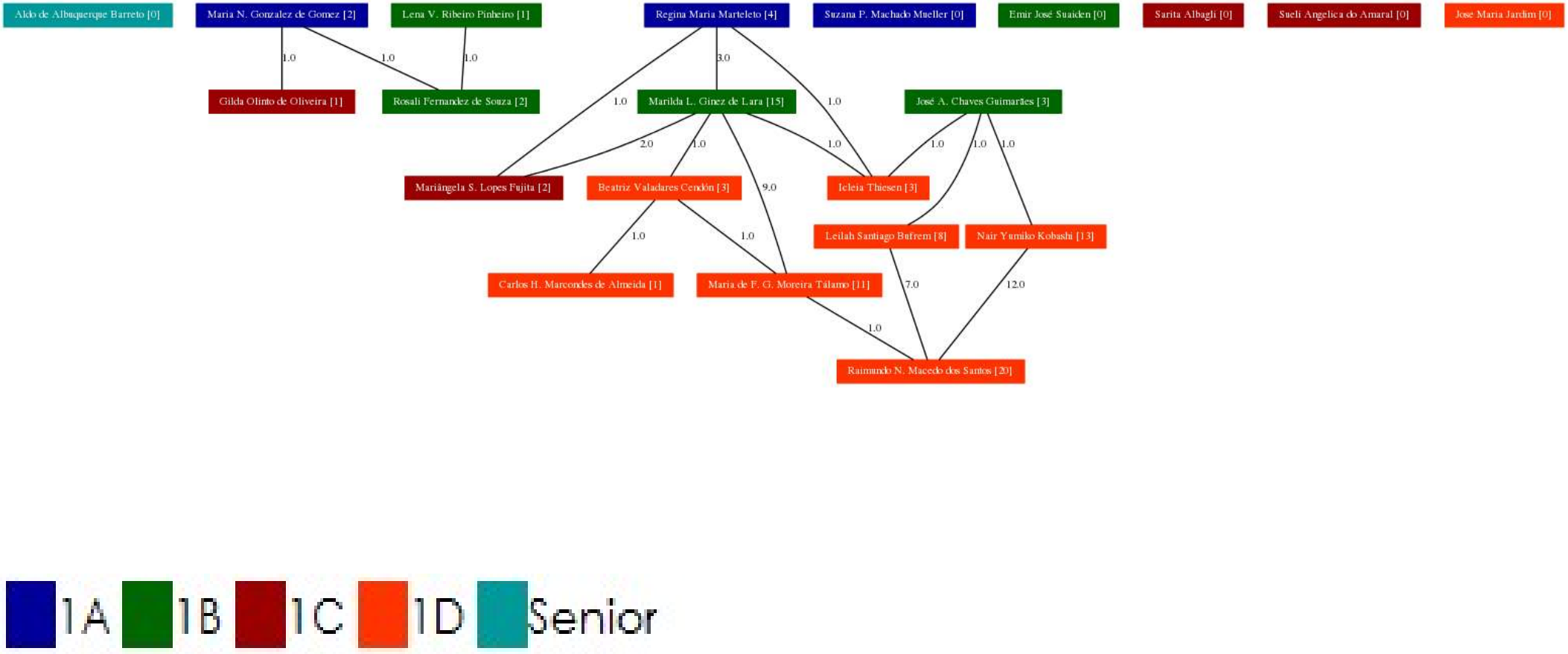


Gráfico 3 – Colaboração entre os Pesquisadores de Produtividade em Pesquisa PQ1 e Sênior do CNPq
Fonte: Plataforma Lattes (Julho, 2015)

A análise das subredes sugere alto grau de colaboração entre os pesquisadores de uma mesma instituição, como é o caso da representação do IBICT no Gráfico 4, cujas linhas evidenciam os graus de contribuição relativos à produção gerada no período delimitador de recorte do corpus.

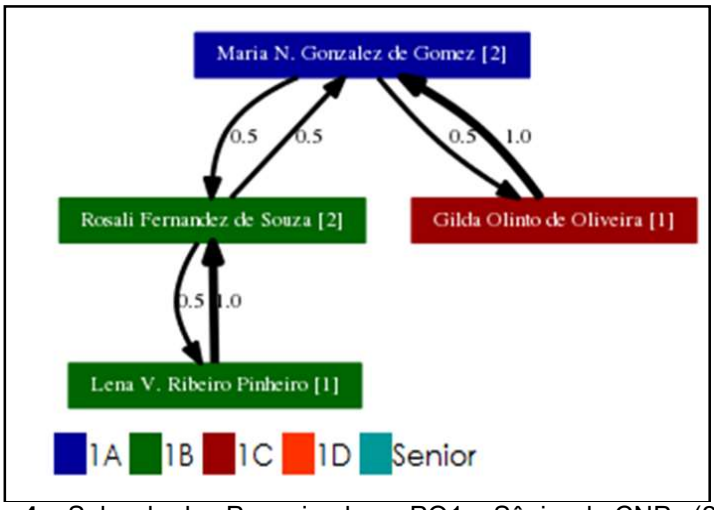


Gráfico 4 – Subrede dos Pesquisadores PQ1 e Sênior do CNPq (2015)
Fonte: Plataforma Lattes (Julho, 2015)

Uma configuração mais aberta e diversificada dessa colaboração é evidenciada no Gráfico 5, ilustrativo do conjunto de pesquisadores das demais instituições representadas no corpus, com seus respectivos graus de intensidade.

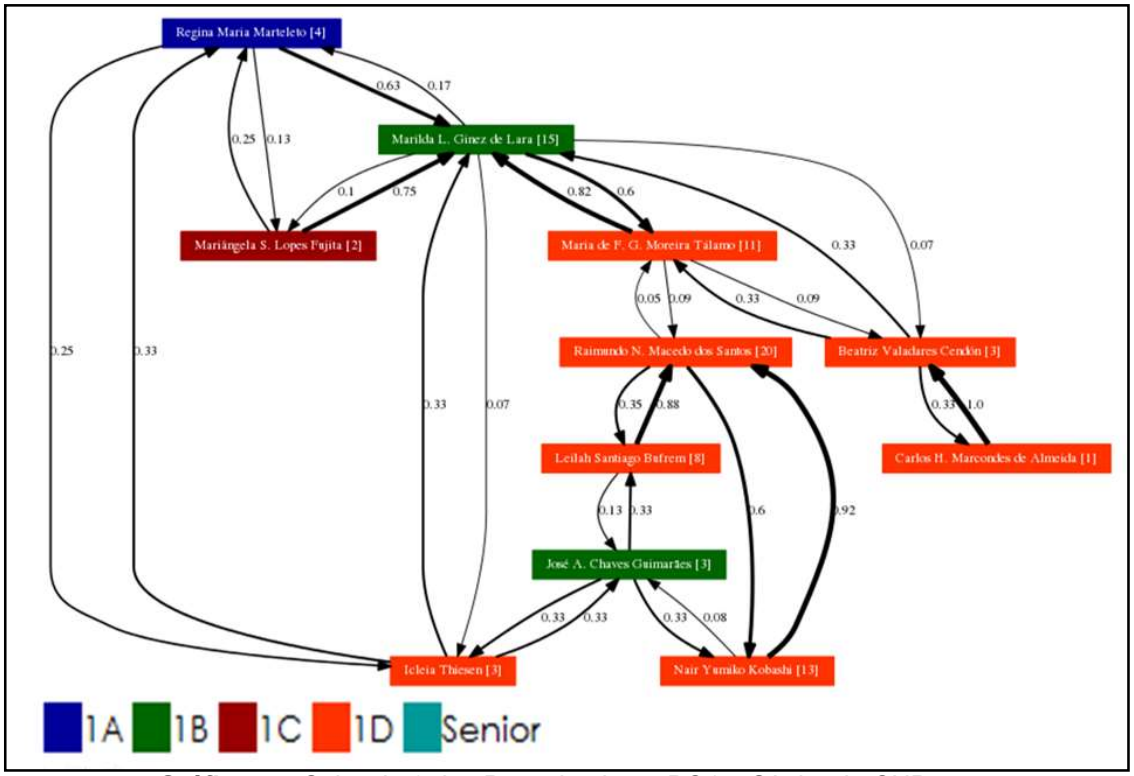


Gráfico 5 – Subrede 2 dos Pesquisadores PQ1 e Sênior do CNPq
Fonte: Plataforma Lattes (Julho, 2015)

Os artigos coletados na Plataforma Lattes foram classificados conforme o estrato Qualis da CAPES dos periódicos em que foram publicados. Como se observa no Gráfico 6, há expressiva incidência de artigos (231) em periódicos classificados como B1, superando a quantidade dos demais artigos publicados em periódicos de outros estratos.

Também chamou a atenção a alta incidência de artigos estratificados no Qualis B3, B4, B5 e C (67), embora possam estar categorizados em outras áreas do conhecimento em um estrato mais elevado.

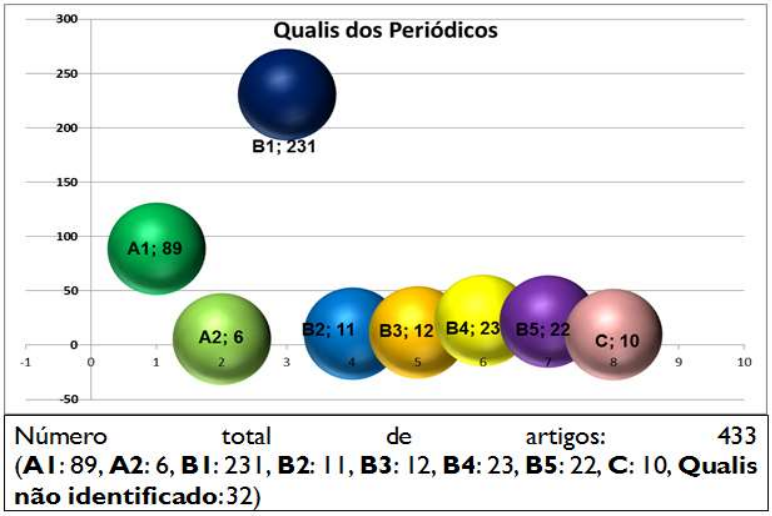


Gráfico 6 – Qualificação da Produção dos Pesquisadores PQ1 e Sênior do CNPq

Fonte: WebQualis (Julho, 2015)

Na **segunda etapa,** entre 7 e 19 de julho de 2015, levantaram-se os artigos científicos indexados na Brapci publicados pelos PQ1 Sênior, A, B, C e D do CNPq. A pesquisa foi feita pelo nome do pesquisador no campo autor, com o recorte de tempo de 2005 a 2015, em revistas científicas.

Após o levantamento das obras de cada um dos autores, foram selecionados apenas os artigos científicos, eliminando-se outros tipos de textos. A análise seletiva determinou que somente os artigos científicos de cada pesquisador fossem exportados para o gerenciador bibliográfico Zotero, utilizado para coletar, armazenar e organizar os registros recuperados. Os artigos foram organizados em coleções que receberam, cada uma, o nome do pesquisador, dentro do Zotero.

Desta seleção, foram recuperados 258 artigos, sendo que oito deles eram de coautoria de dois pesquisadores do grupo de PQ1. A inclusão do Gráfico 7 foi motivada pela possibilidade de se observar um retrato da produção total dos autores,

em que a configuração temática é mais ampla do que no gráfico 8, devido ao recorte específico de interesse do Congresso ISKO Brasil 2015.

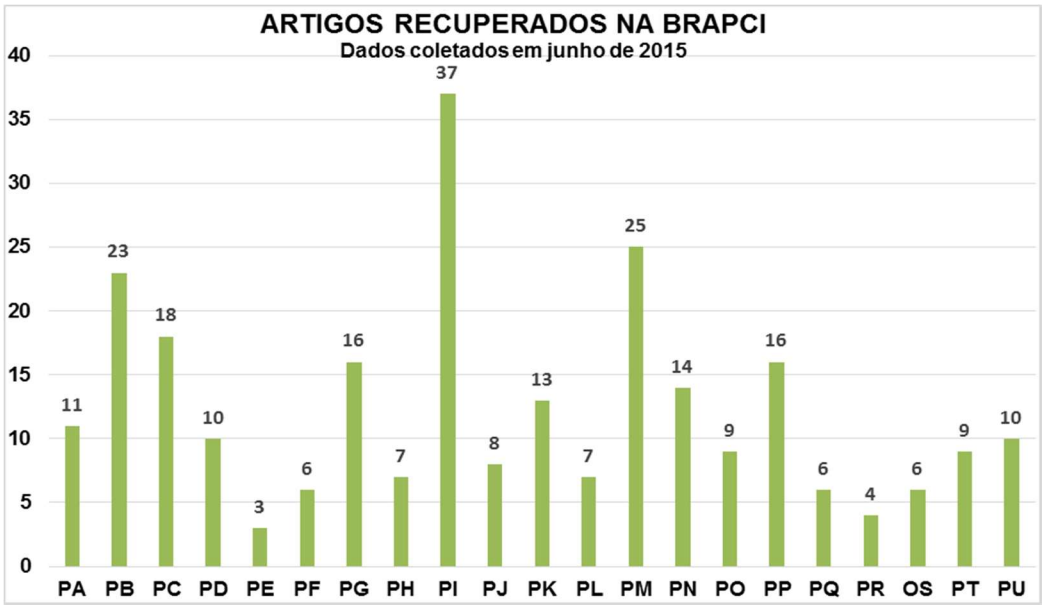


Gráfico 7 – Artigos dos Pesquisadores PQ1 e Sênior, recuperados na Brapci, de 2006 a 2015
Fonte: Brapci (2015)

Na **terceira etapa,** realizou-se a leitura dos títulos, resumos e palavras-chave dos 250 artigos para que fossem identificados e selecionados como componentes do corpus apenas aqueles que se relacionassem à temática OC, totalizando 104 artigos, ou seja, 41,6% do total de toda a produção dos PQs, o que se pode considerar expressivo. Estes foram categorizados conforme as três dimensões que norteiam as discussões do III ISKO Brasil: 1) a **dimensão epistemológica** da organização do conhecimento, incluindo as bases conceituais, históricas e metodológicas da OC, assim como os diálogos interdisciplinares que se estabelecem em seu âmbito; 2) a **dimensão aplicada** da OC, nela incluídos os modelos e formatos, os instrumentos, os produtos e as estruturas em OC e 3) a **dimensão social e política** da OC, abrangendo temas relativos à atuação profissional, à ética, à cultura e à identidade, assim como ao desenvolvimento sustentável. Assim categorizados, os artigos foram analisados e interpretados e mapeados conforme as subcategorias de cada dimensão, conforme explicitado no próximo capítulo.

4 Apresentação e análise dos dados

Os 104 artigos relacionados ao domínio de OC, no período de 2005 a 2015, publicados em periódicos científicos indexados na BRAPCI, estão distribuídos, no Gráfico 8, entre os pesquisadores cujos nomes foram substituídos por letras para que

se preservasse o anonimato. Por terem sido escritos em coautoria entre os pesquisadores PQ1, três dos artigos apresentam duas ocorrências.

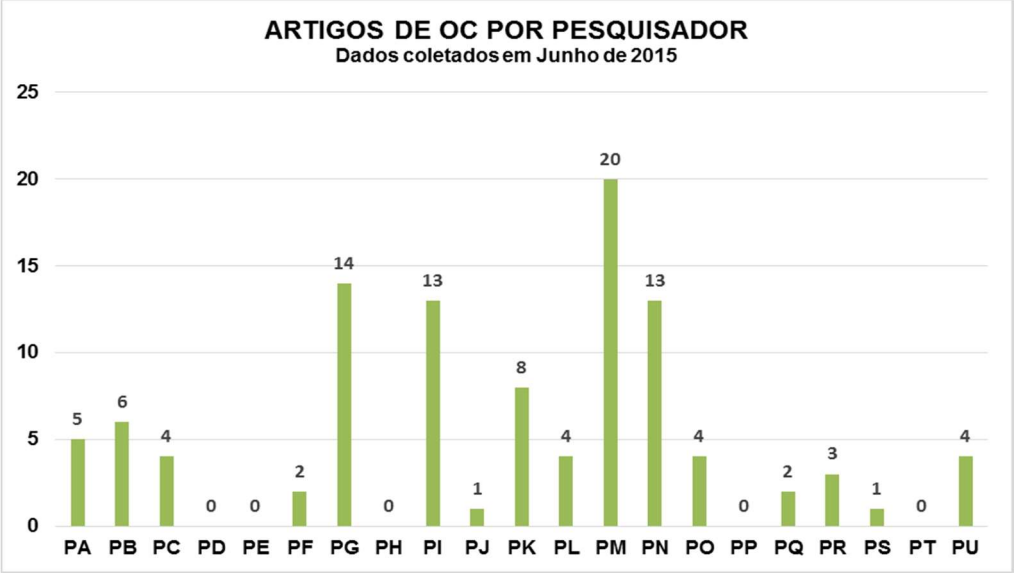


Gráfico 8 – Artigos dos Pesquisadores PQ1 e Sênior, recuperados na Brapci, de 2006 a 2015

Fonte: Elaboração Própria (2015)

Os periódicos mais representados no corpus específico de OC foram as revistas DataGramaZero, com 20 artigos, a Encontros Bibli, com 15 artigos e a Transinformação com 12 (Gráfico 9). A maioria (13) delas está classificada no estrado B1 do Qualis CAPES, conforme o Gráfico 9. Percebe-se que, embora não sejam revistas com muitos anos de existência, as duas primeiras têm sido mais receptivas aos temas da OC.

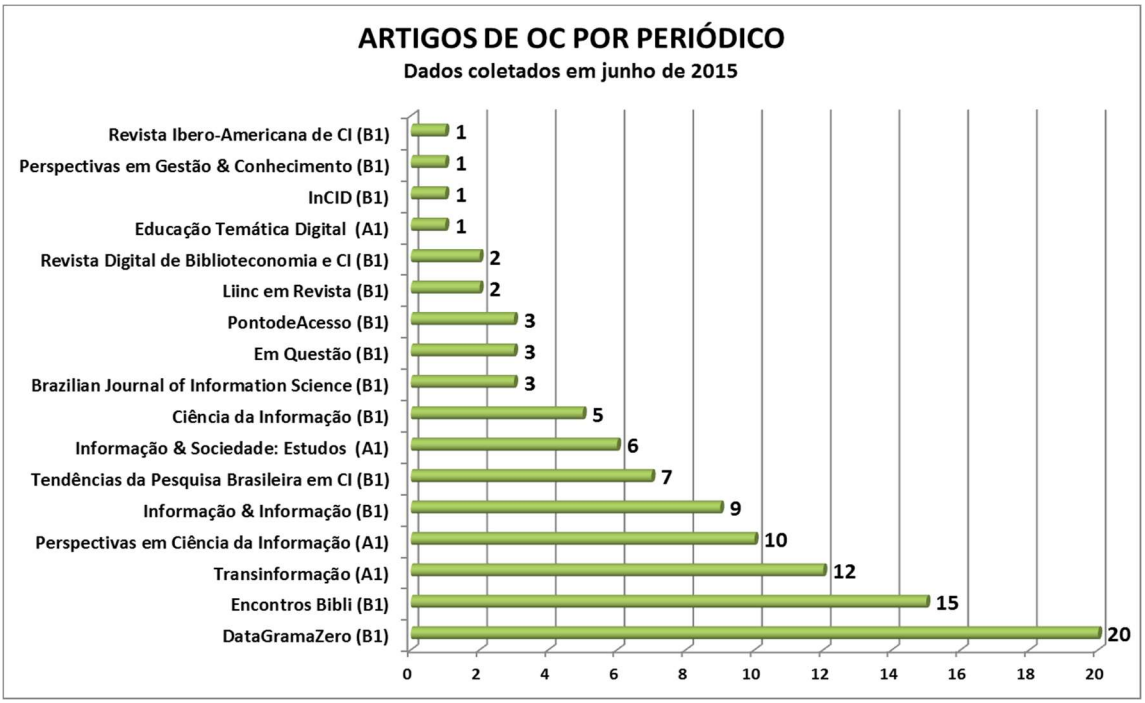

ARTIGOS DE OC POR PERIÓDICO
Dados coletados em junho de 2015
Revista Ibero-Americana de CI (B1) 1
Perspectivas em Gestão & Conhecimento (B1) 1
InCID (B1) 1
Educação Temática Digital (A1) 1
Revista Digital de Biblioteconomia e CI (B1) 2
Liinc em Revista (B1) 2
PontodeAcesso (B1) 3
Em Questão (B1) 3
Brazilian Journal of Information Science (B1) 3
Ciência da Informação (B1) 5
Informação & Sociedade: Estudos (A1) 6
Tendências da Pesquisa Brasileira em CI (B1) 7
Informação & Informação (B1) 9
Perspectivas em Ciência da Informação (A1) 10
Transinformação (A1) 12
Encontros Bibli (B1) 15
DataGramaZero (B1) 20
0 2 4 6 8 10 12 14 16 18 20

Gráfico 9 – Artigos de OC por periódico, recuperados na Brapci, de 2006 a 2015
Fonte: Elaboração Própria (2015)

Os 101 artigos de OC foram categorizados cada um em uma das três dimensões: Epistemológica; Aplicada; Social e Política. Percebe-se maior expressividade quantitativa de artigos na Dimensão Aplicada (45 artigos), enquanto na Dimensão Epistemológica foram categorizados 41 artigos e na Dimensão Social e Política 15 artigos (Gráfico 10).

Inferiu-se, a partir da análise de conteúdo que, entre os artigos categorizados na *dimensão aplicada* há predominância das pesquisas voltadas aos modelos, formatos, instrumentos, produtos e estruturas de organização do conhecimento. Entre elas, um grupo específico, o mais representativo nos estudos, relaciona-se aos processos de indexação, sob responsabilidade quase exclusiva de PM. A categoria inclui artigos sobre os processos de indexação, de sistemas de informação e formas de representação documentária na indexação (modelo de leitura documentária para textos científicos), protocolo verbal e avaliação da indexação.

As estruturas de organização do conhecimento nesse grupo de artigos incluem, além dos modelos apresentados e testados, as tarefas bibliotecárias e contextualizadas em bibliotecas e unidades de informação, observadas diante da situação atual, marcada por transformações tecnológicas significativas. Essas tarefas relacionam-se a temas como indexação de livros, linguagem documentária, consistência da indexação, indexação na catalogação de bibliotecas universitárias,

indexação automática de artigos de periódicos, utilizações e avaliações de softwares. Também são tratadas questões relativas às linguagens documentárias, interoperabilidade e vocabulário controlado, às relações entre linguística e terminologia, sistemas classificatórios e princípios de agregação.

Entre os produtos de organização do conhecimento são analisadas e avaliadas linguagens documentárias, incluindo-se os índices e tesauros e os sistemas classificatórios.

Ainda como subsídios às estruturas de OC, são estudados as metodologias de avaliação, os princípios de agregação segundo contextos e áreas do conhecimento, a metodologia de representação da informação imagética, a garantia literária, a eficácia probatória dos documentos e a aplicação prática da análise de conteúdo e de citações para análise e recuperação de documentos.

A contribuição para o aperfeiçoamento das linguagens documentárias tem sido uma constante nos trabalhos dos pesquisadores de PM, PN, PG e PO, cujas motivações envolvem as questões relacionadas à avaliação de linguagens pelo uso, incluindo estudos sobre linguagens natural e controlada e pesquisas sobre indexação sistemática. Vocabulários controlados são analisados em trabalhos como os de PM e PN, que também se fundamentam em avaliações quantitativas de linguagens documentárias, alfabéticas e hierárquicas.

Observam-se ainda temas emergentes relacionados às estruturas e modelos de OC, tais como a intensificação dos processos de interação humana, as representações nos modos de apresentação da literatura, os métodos de avaliação na organização da produção científica e a criação automática de bases de citações, os portais e sua representatividade.

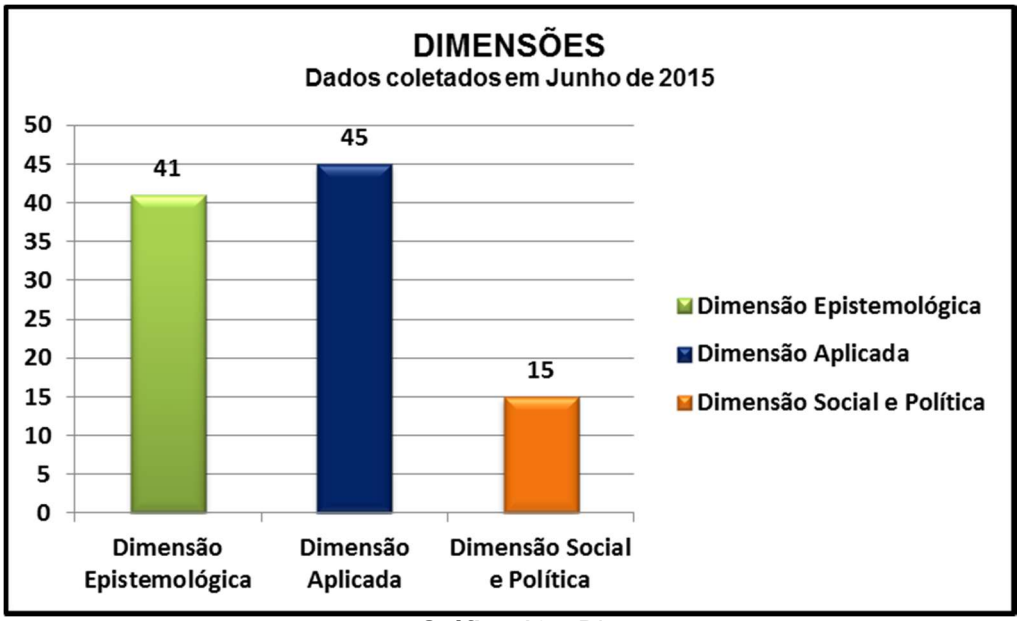


Gráfico 10 – Dimensões

Fonte: Elaboração Própria (2015)

A *dimensão epistemológica*, representada principalmente por PG, PI, PK e PN, concretiza-se principalmente em relação aos diálogos interdisciplinares na OC, tema dominante em estudos sobre o referencial teórico-metodológico do domínio. Por sua vez, estudos sobre as bases conceituais do conhecimento refletem um amplo espectro teórico.

As bases metodológicas são representadas por artigos predominantemente voltados à relação entre o domínio da organização do conhecimento e as práticas pedagógicas.

Quanto às bases cognitivas, são o tema de artigos que, de modo especial, tratam do exame do texto na formação do conhecimento, assim como do processo dialógico entre o processo de conhecer e o contexto social dos sujeitos.

A *dimensão social e política*, embora menos representada quantitativamente, exerce, ao lado da dimensão epistemológica, papel significativo enquanto favorece o equilíbrio em relação à dimensão aplicada. Representada expressivamente por artigos que se voltam à problemática correspondente à formação e à atuação profissional, essa categoria teve a participação efetiva de PM. Foi secundada pelos artigos que tratam das questões éticas em OC, dimensão representada especialmente por PG. Quanto às questões relativas a cultura e identidade em relação à OC, pautam-se de modo difuso, tangenciando temas como gênero e ética.

5 Considerações finais

As evidências, a partir da análise voltada à expressividade quantitativa das ocorrências, iniciam pela visibilidade de estudos relativos à dimensão aplicada, destacando-se o papel da OC no universo da CI e, de certo modo, corroborando a posição nuclear do Tratamento Temático da Informação (TTI), destacada por Guimarães (2008), e justificada por sua natureza mediadora entre a produção e o uso da informação. Essa “nuclearidade” está relacionada ao fato de que somente cinco pesquisadores da área de CI não fazem parte do universo de 21 participantes do universo do pesquisadores Sênior e A1.

A evidência quantitativa relativa à dimensão aplicada, portanto pragmática e voltada a realidades concretas, é harmonizada pela dimensão epistemológica, também com expressiva representação em estudos relativos aos fundamentos teórico-conceituais, históricos e tecnológicos do domínio, assim como sobre as relações interdisciplinares.

E como a comprovar a impossibilidade de se compartimentalizarem as visões de uma conjuntura dada, evidenciaram-se nas pesquisas integrantes da dimensão política e social, tanto reflexões de caráter epistemológico quanto análises sobre questões aplicadas, com distintas ênfases.

Sem prejuízo do aprofundamento futuro sobre os aspectos evidenciados, pode-se confirmar a significativa diversidade dos estudos e a representatividade orgânica de pesquisadores distinguidos no seu campo de atuação.

Do exposto, destaca-se a oportunidade da análise de conteúdo dos artigos, pois ela revelou a OC como um domínio não somente apoiado em estudos aplicados, voltados às necessidades e desafios do seu cotidiano, mas também inspirado no compromisso social, cuja missão vem sendo cumprida ao amparo de princípios e teorias fundamentais para as decisões relativas aos modos de organizar o conhecimento. Revelou, também, aproximações significativas entre as pesquisas registradas nos periódicos e os projetos originais dos pesquisadores, graças às referências ao órgão de fomento nas publicações. Por outro lado, as coautorias destacam-se na diversidade dessa produção, pois se verificam, na grande maioria dos casos, aliando orientandos e orientadores. Entretanto, essa aparente dispersão relativa ao foco das pesquisas registradas, se percebida nas relações que se

produzem na cultura da pós-graduação, tem revelado o vínculo necessário entre os projetos do pesquisador e dos orientandos.

Referências

BARITÉ, M. Organización del conocimiento: un nuevo marco teórico-conceptual en Bibliotecología y Documentación. In: CARRARA, K. (org.). **Educação, Universidade e Pesquisa**. Marília: Unesp-Marília-Publicações, São Paulo: Fapesp, 2001. p. 35-60.

BORTOLOZZI, F.; GREMSKI, W. Pesquisa e pós-graduação brasileira–assimetrias. **Revista Brasileira de Pós-Graduação**, v. 1, n. 2, 2011.

BOURDIEU, P. Le champ scientifique. **Actes de la Recherche en Sciences Sociales**, n. 2/3, jun. 1976, p. 88-104.

BOURDIEU, P.; PASSERON, J.-C. **A reprodução**: elementos para uma teoria do sistema de ensino. Petrópolis: Vozes, 2014.

ESTEBAN NAVARRO, M. A.; GARCÍA MARCO, F. J. Las primeras jornadas sobre organización del conocimiento: organización del conocimiento e información científica. **Scire**, v.1, n.1, p.149-157, jan./jun. 1995.

GUIMARÃES, J. A. C. A dimensão teórica do tratamento temático da informação e suas interlocuções com o universo científico da International Society for Knowledge Organization (ISKO). **Revista Ibero-americana de Ciência da Informação (RICI)**, Brasília, v. 1, n. 1, p. 77-99, 2008.

HJØRLAND, B. What is Knowledge Organization? **Knowledge Organization**, v. 35, n.2/3 2008.

LLOYD, C. **As estruturas da história**. Rio de Janeiro: Zahar, 1995. 296 p.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal Of The Brazilian Computer Society,** Porto Alegre, v. 4, n. 15, p.31-39, 2009. Disponível em: <http://www.scielo.br/pdf/jbcos/v15n4/04.pdf>. Acesso em: 5 jun. 2015.

MENDES, P. H. C. et al. Perfil dos pesquisadores bolsistas de produtividade científica em medicina no CNPq, Brasil. **Revista Brasileira de Educação Médica**, Rio de Janeiro, v. 34, n. 4, dez. 2010. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-55022010000400008&lng=en&nrm=iso>. Acesso em: 09 jun. 2015. http://dx.doi.org/10.1590/S0100-55022010000400008.

SANTOS, A. L. F. dos; AZEVEDO, J. M. L. de. A pós-graduação no Brasil, a pesquisa em educação e os estudos sobre a política educacional: os contornos da constituição de um campo acadêmico. **Revista Brasileira de Educação**, v. 14, n. 42, p. 535, 2009.

WASSERMAN, S.; FAUST, K. **Social network analysis**: methods and applications. Cambridge: Cambridge University Press, 1994. 857 p.

Pesquisa em análise documentária no Brasil: a influência de Jean-Claude Gardin

A busca pela eficiência na representação da informação e do conhecimento – desdobramentos posteriores no pensamento de Gardin

Johanna W. Smit
Universidade de São Paulo
johannawsmit@gmail.com

1 Introdução

A relação entre o texto original e sua representação acompanhou Gardin ao longo de toda sua trajetória, e ele a discutiu não somente pensando em sistemas de informação, mas nas diferentes modalidades de análise de textos.

Este não é o momento para detalhar uma bibliografia importante na qual Gardin discutiu os diferentes métodos de análise de textos, mas uma rápida referência à questão não pode ser ignorada, tendo em vista que a análise documentária (AD), em sua visão, representa uma modalidade de análise de textos – segundo ele, o ramo mais "industrial" das análises (Gardin, 1974, p.45). Enfatizando a dimensão semântica das análises, ele elaborou uma visão comparativa dos procedimentos adotados, por exemplo, pela análise de conteúdo exercida por sociólogos e outros profissionais, a análise estrutural de contos e narrativas, a análise semiológica ou literária e a análise documentária, ou seja, "a expressão do conteúdo de textos científicos à maneira dos documentalistas" (Gardin, 1970, p.630). Estas comparações o levaram a postular a importância da explicitação do método de análise de textos (em sua formulação: como passar de T1, original, a T2, sua representação), a explicitação das correspondências semânticas mobilizadas pela análise e a possibilidade de verificação da relação entre o texto original (T1) e sua representação (T2). Dito em outros termos, ele propôs discussões acerca dos **métodos de análise** (e sua explicitação), das **ferramentas** (ou seja, a metalinguagem usada para extrair o conteúdo semântico dos textos, de

acordo com os objetivos propostos) e estratégias de **avaliação** e **validação** das análises.

Inserir a análise documentária num conjunto maior de análises de textos representou, certamente, uma ampliação do conhecimento na documentação e biblioteconomia, ao chamar a atenção para questões – linguísticas - até então consideradas resolvidas pelo bom-senso. Entre a atribuição de um código de classificação a um livro e as discussões sobre a unidade de análise (palavra? frase? parágrafo? capítulo? o texto inteiro?), a preponderância da semântica sobre a sintaxe e o potencial de automatização do processo de classificação, elaboração de resumos ou indexação separam o bom-senso de uma visão que se quer mais científica e rigorosa da prática biblioteconômica ou documentária.

A importância da metalinguagem (ML) usada para analisar os textos - a linguagem documentária (LD) na Ciência da Informação - recebeu muita atenção e suponho que esta questão seja a que mais foi incorporada pela Ciência da Informação. Uma linguagem para representar o conteúdo original dos textos deve ser uma linguagem artificial, construída de acordo com objetivos e, portanto, distinta da linguagem natural. A origem da temática pode ser encontrada quando o arqueólogo Gardin discutiu como encontrar correspondências (ou seja, a classificação) entre objetos arqueológicos, ou seja, quais critérios adotar para reunir objetos em conjuntos, distinguindo-os de outros conjuntos. Nas décadas de 50 e 60 ele elabora “códigos analíticos” de cilindros orientais, artefatos, desenhos decorativos, etc. enfatizando sempre a necessidade de comparações para estabelecer relações (Gardin, 1979, p.87).

A preocupação com a importância do estabelecimento de relações entre dados e informações fica clara também, quando ele faz parte da equipe que discute a concepção do UNISIST na Unesco, em 1970, ao afirmar que “não adianta disponibilizar os dados, as pessoas têm que conseguir entender os dados, estabelecendo as relações entre os dados”.

É verdade que a tecnologia da época nos parece hoje bem rudimentar, pois a seleção de objetos que apresentavam características comuns foi concebida em fichas com perfuração lateral (peek-a-boo). Uma análise dos conceitos presentes no Alcorão também deu margem à elaboração de um vocabulário, estruturado, de conceitos, e testes feitos com estas fichas (Allard et al, 1963).

O investimento em linguística foi muito grande (Gardin, 1973): consideramos que hoje o Grupo TEMMA deu prosseguimento a este investimento, mas com ênfase na terminologia, visando questões relacionadas ao controle de vocabulário e o poder comunicacional dos produtos da análise documentária.

A necessidade de teorização das linguagens documentárias, ou seja, as metalinguagens específicas dos ambientes de tratamento da informação, levou Gardin a discutir suas estruturas e sua composição, distinguindo nelas um **léxico, relações paradigmáticas** e **relações sintagmáticas** (Gardin, 1966, 1970, 1973). Ou seja, aquilo que até aquele momento muitos faziam baseados no bom-senso, passou a contar com uma teoria, fortemente ancorada na linguística.

Em relação ao **léxico**, ou seja, os termos que compõem a linguagem documentária, deve-se frisar a necessidade de seu controle, eliminando a sinonímia, resolvendo a homonímia, etc.

Em relação às **relações paradigmáticas -** principalmente as relações hierárquicas reconhecidas e estabelecidas pelos diferentes domínios do conhecimento - seu maior interesse reside na afirmação de que as mesmas não representam automaticamente a organização de determinado universo do conhecimento, mas os objetivos do serviço de informação, o que reforça o caráter artificial das linguagens documentárias.

No entanto, desde a década de 60, Gardin se preocupava com as relações que podem ser estabelecidas entre termos do léxico, mas que não configuram uma relação paradigmática, ou seja, as **relações sintagmáticas**, chegando ao ponto de codificá-las quando do desenvolvimento do SYNTOL (Syntagmatic Organization Language), como será detalhado a seguir.

A otimização da relação entre informação original (T1) e informação representada (T2) pode ser analisada, nas propostas de Gardin, por três desdobramentos:

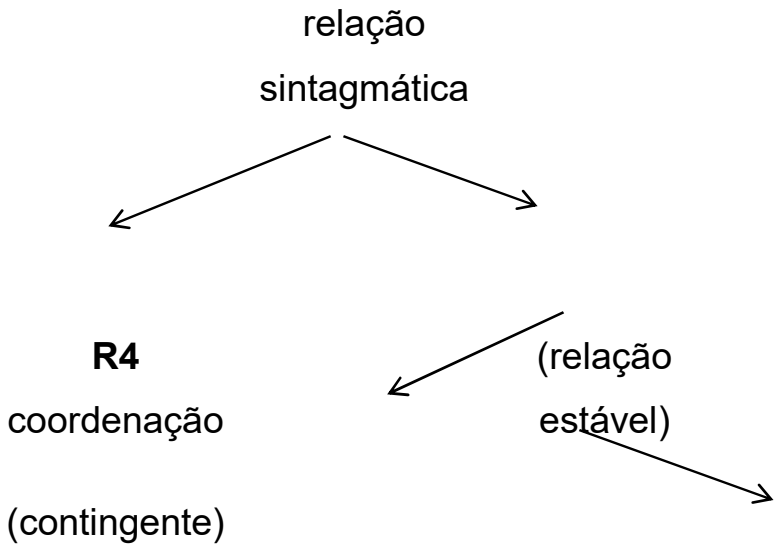
1. Propor a construção de ontologias (cf. seção 1);
2. Adotar uma visão extremamente positivista da ciência (cf. seção 2) e
3. Face ao custo envolvido nas propostas que serão apresentadas e discutidas nas seções 1 e 2, ele imaginou outra abordagem, mais radical, e que propunha a reformulação do texto científico, visando torná-lo mais consultável (cf. seção 3).

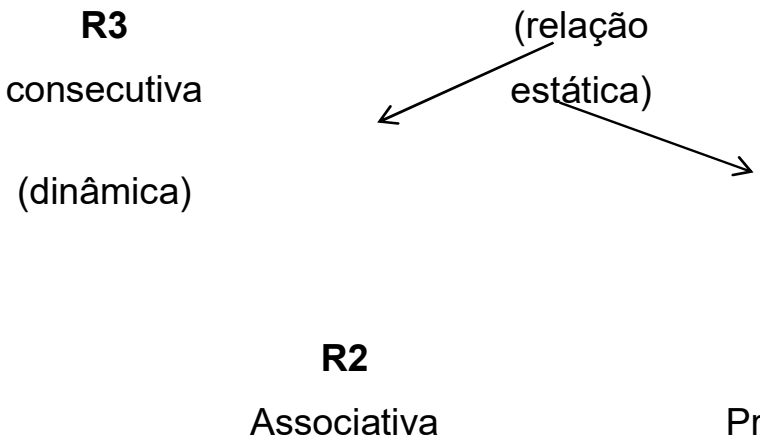
1 A proposta de ontologias antes da hora

O SYNTOL foi desenvolvido na França, por um grupo coordenado por Gardin, por encomenda da Euratom (European Atomic Energy Community) entre 1960 e 1962. O SYNTOL não é uma LD mas um sistema geral de documentação automática, objetivando a indexação automática de textos, composto de um léxico organizado (ou seja, termos controlados e organizados paradigmaticamente) e um sistema bastante elaborado de relações sintagmáticas, para estabelecer a interconexão entre os termos no caso de assuntos mais complexos. "O SYNTOL é um conjunto de regras lógico-linguísticas, a respeito das diferentes maneiras de representar as informações que se encontram nos documentos científicos, com vistas a sua exploração pelo computador" (Cros, Gardin, Levy, 1964, p.39-40).

Extremamente detalhado e muito complexo, a concepção do SYNTOL é muito interessante do ponto de vista teórico, mas pouco usado na prática. Testes foram realizados em 1962, para textos das áreas de fisiologia, psicologia, sociologia e etnologia e permitiram concluir que o Syntol, nas palavras de Gardin, era "muito canhão para pouca mosca".

Se a implantação prática ficou comprometidao (por conta tanto da complexidade do sistema como das limitações tecnológicas da época), importa reter da experiência o caráter precursor da proposta de padronização de relações entre os termos no momento da indexação. As relações sintagmáticas foram organizadas e codificadas em conjuntos lógicos (donde a expressão OLS - Organização Lógico-Semântica) e apresentadas na forma de diferentes árvores de relações. Por exemplo:





A relação entre “álcool” e “alegria”, por exemplo, produzia o sintagma <álcool R3 alegria>, onde R3 denotava uma relação consecutiva. Deve-se salientar que entre as décadas de 70 e 90 o recurso aos *role operators,* ou marcadores, foi muito utilizado. Trabalhando com textos na área médica, por exemplo, passava a ser importante poder distinguir uma droga na condição de medicamento para combater certa doença ou na condição de provocador de uma doença. Hoje parece simples associar o que foi proposto àquela época às atuais ontologias, mas seu caráter precursor, com uma tecnologia ainda incipiente, não pode ser ignorado.

A formalização das relações levou à definição do sintagma mínimo de qualquer representação da informação: Ri (x,y), onde *R* representa uma relação (paradigmática ou sintagmática), *x* e *y* são termos do léxico. A partir da multiplicação de LDs e da elaboração de tesauros multilíngues, a noção do sintagma mínimo mostrou seu potencial nos estudos de compatibilidade e interoperabilidade entre LDs.

Deve-se reter destas propostas, datada da década de 60, a evidente conexão com os princípios da inteligência artificial, também em amplo desenvolvimento naquela época, através da explicitação de relações entre os termos e análise de condições nas quais as relações podem ser estabelecidas. A mesma temática é hoje atualizada no desenvolvimento de ontologias, ao identificar “objetos”, propriedades, regras de relacionamentos e condições de aplicabilidade: este conjunto compõe a base de conhecimento e o motor de inferência. (Gardin, 1989, p.16-17).

2 A avaliação da representação da informação

A capacidade representacional dos produtos da AD recebeu muito atenção por parte de Gardin e certamente representa um veio de pesquisa ainda pouco explorado pelo Grupo TEMMA. Esta avaliação do produto da AD pode ser sistematizada a partir de dois pontos de vista, complementares mas totalmente distintos:

- a avaliação da "satisfação" do usuário a partir de uma busca em sistema de informação. Cálculos de precisão, revocação, ruído, etc. foram muito desenvolvidos pela bibliografia e são, evidentemente, muito importantes;
- mas Gardin considerava outro cálculo igualmente importante e que dizia respeito ao grau de representatividade do produto da análise em relação aos textos originais (Gardin et al, 1981).

As duas avaliações não devem ser consideradas auto-excludentes, mas complementares, embora a segunda avaliação, que enfatiza a representatividade da representação em relação ao texto original não seja comum em nossa bibliografia.

Sobrepõe-se, nesta questão, a meu ver, a figura do cientista àquela do documentalista. É interessante observar que a representatividade do texto 2, produzido a partir do texto 1 é na prática, sempre pressuposta, mas não discutida. Gardin não transformava o tema em uma questão ética ou moral mas numa questão prática: ele considerava que o usuário do sistema de informação para poder, efetivamente, avaliar o resultado de suas buscas no sistema de informação, devia ter acesso a toda informação relacionada aos procedimentos de AD adotados, bem como todos os detalhes da LD empregada. Somente o conhecimento de todos os detalhes que integram o processamento da informação faz com que o usuário possa sair de uma posição passiva – receber o resultado de uma busca – para uma posição ativa de real avaliação do resultado da busca efetuada.

Um usuário em condição de avaliar, autonomamente, o resultado de uma busca e, portanto, validá-lo: este é um desafio lançado por Gardin que, ao que me parece, ainda está longe de ser respondido. Mas o desafio é bonito, instigador, e merece integrar a agenda de pesquisa. A satisfação do usuário, neste caso, se completa por uma recuperação de informações que não somente corresponde aos termos utilizados em sua pesquisa mas que remetem a textos que igualmente correspondem ao que foi buscado.

3 O projeto logicista e a re-escrita do texto científico

Naquilo que me parece ser a última fase das empreitadas de Gardin, ele se interroga cada vez mais sobre o sentido social das publicações científicas e dos sistemas de informação. Na década de 80 e parte da década de 90 ele investiu muito na temática da análise de textos, em diferentes áreas, tentando identificar a consistência das mesmas: ao final ele considerou demonstrado que os textos nas áreas das ciências sociais e humanas apresentavam muitos vícios, recheavam publicações e bibliotecas mas que não era possível perceber o avanço do conhecimento. Obviamente esta conclusão é discutível, mas faz sentido quando colocada na perspectiva gardiniana da total explicitação dos procedimentos de representação da informação. Inconformado com este estado de coisas, ele elabora novas propostas, alicerçadas em todo o trabalho anterior, objetivando uma solução mais lógica, ou útil, para a produção do conhecimento em ciências sociais e humanas.

Reaproximando-se novamente da questão da documentação, ele retoma, sem citar, a dualidade enunciada por Briet, ao identificar uma ordem primária dos textos, produto da atividade científica e técnica que é apresentada de acordo com hábitos estabelecidos pelas diferentes áreas do conhecimento e uma ordem secundária, na qual ocorre a produção da informação documentária e criação de acessos à ordem primária. Gardin considera que a documentação se tornou refém da ordem primária, tendo em vista que a produção de textos científicos aumentou exponencialmente.

Em seguida, ele postula estar ocorrendo uma “crise das publicações”, em função de um desequilíbrio flagrante entre a produção de textos científicos e a capacidade humana de consumi-los. A criação de ferramentas para acesso aos documentos originais - pela melhoria de procedimentos de seleção de textos através da indexação, elaboração de resumos, sistemas de alerta, mineração de dados e etc. - evidentemente amenizam o desequilíbrio entre produção e consumo, mas não resolvem o problema, até porque a capacidade de consumo de informação, pela mente humana, mudou muito pouco nos últimos séculos, face ao crescimento exponencial da publicização de textos. “A busca incessante de otimização entre silêncios e ruídos” permanece, mesmo quando amenizada (Gardin, 2001, p.2).

A próxima pergunta que Gardin formula é mais provocante: em supondo um sistema de documentação totalmente adequado, que não elimine documentos pertinentes em resposta a uma consulta, aliado a uma seleção bem calibrada de textos pertinentes, pode-se postular que todos deveriam ser lidos, pois pertinentes, mas que

certamente somente uma parte será lida integralmente e que muitos serão simplesmente consultados. Ele cita experimentos feitos de seleção de textos em determinada área, por especialistas para atestar sua pertinência, e cálculos do tempo necessário para sua leitura, concluindo pela total impossibilidade de leitura de todos os textos considerados pertinentes. A partir desta constatação, Gardin propõe a remodelagem dos textos científicos de forma a que possa sem consultados, tendo em vista que não serão lidos. A “crise das publicações científicas” levanta ainda outra questão: quantos leitores, em média, terá um artigo científico? “É razoável manter um sistema de publicação de artigos cujos leitores são tão raros”? (Gardin, 2001, p.3).

A partir destas interrogações Gardin estima que o problema não está em tornar acessível os melhores textos (isto a documentação sabe fazer e faz muito bem), mas em conseguir consumi-los. Pode-se, portanto, reformular a afirmação de Bush, quando este propunha “a tarefa massiva de tornar mais acessível um desconcertante acervo de conhecimentos” (Bush, 1945, p.1) e afirmar que não é mais possível continuar escrevendo textos para serem lidos, quando a maior parte será somente consultada. Enquanto os textos científicos não forem remodelados, a documentação continuará refém do sistema de publicações científicas.

Trata-se, portanto, de uma proposta de intervenção na produção de publicações, decompondo-as em diferentes camadas:

- fatos e dados – vestígios materiais, acontecimentos históricos, práticas culturais;
- teorias, ou seja, pontos de vista formulados sobre os fatos ou dados;
- os procedimentos adotados para passar dos fatos às interpretações e teorias;
- os exemplos citados para corroborar a elaboração de interpretações e teorias.
- Percebe-se nesta proposta a volta dos conceitos da base de conhecimento e do motor de inferência, caros aos sistemas especialistas. As camadas, hipertextuais, devem, em sua visão (Gardin, 2001):
- preservar o conteúdo informacional dos textos;
- explicitar a base de conhecimento e as regras de passagem de T1 para T2 (regras de raciocínio);

- excluir as digressões metodológicas, recursos retóricos, referências filosóficas, charmes literários, etc.

A proposta foi concretizada ao remodelar um livro de 500 páginas - sobre as pérolas de cornalina, recolhidas em sítios arqueológicos da Índia e do Oriente Próximo e que levaram a inferências sobre as sociedades que as fabricavam ou importavam - em um CD-ROM multimídia contendo o mesmo conteúdo informacional, com navegação hipertextual, separando e relacionando fatos, ilustrações, argumentos de apoio, conclusões, esquematizações dos procedimentos adotados para passar dos fatos às conclusões, etc.

A categorização de tipos de informação é seguramente uma competência desenvolvida pela Ciência da Informação, razão pela qual Gardin propõe um trabalho cooperativo entre profissionais de informação, editores e autores, reconhecendo que a clareza acerca de léxicos e relações entre termos é apanágio dos profissionais de informação.

4 Em resumo

Além das contribuições de Gardin ao introduzir a linguística nos estudos de classificação, AD e indexação, restam-nos os desafios por ele lançados:

- reconhecer a documentação como uma etapa linguística do fazer científico, ao estabelecer terminologia e discutir as relações entre termos, de acordo com os campos do conhecimento;
- desenvolver sistemas de avaliação dos produtos de nosso trabalho que aliem os estudos de satisfação dos usuários a avaliações da representatividade de nossos produtos em relação aos textos representados;
- investir nas ontologias como forma de contribuição à indústria de publicações de textos científicos, para torna-los consultáveis. Neste esforço os dados inciais e as regras de inferência devem ser totalmente explicitados: a disponibilização de dados abertos já é muito discutida hoje, mas a explicitação das regras de inferência , que supõem a repetibilidade nas ciências, ainda é um tabu nas ciências sociais e humanas.

Referências

ALLARD, M., ELZIERE, M., GARDIN, J.-C., HOURS, F.(1963). Analyse conceptuelle du Coran sur cartes perforées. Paris: Mouton.

BUSH, V. (1945). As we may think. Atlantic Monthly, julho.

CROS, R.C., GARDIN, J.-C., LEVY, F. (1964). Le Syntol, um système général de documentation automatique. Paris: Gauthier Villars, 1964.

GARDIN, J.-C. (1989). Artificial intelligence and the future of semiotics: an archaeological perspective. Semiotica, vol.77, n.1/3, p.5-26.

GARDIN, J.-C. (1973). Document analysis and linguistic theory. Journal of Documentation, vol. 29, n.2, p.137-168.

GARDIN, J.-C. (1966). Eléments d'um modéle pour la description des lexiques documentaires. Bulletin des Bibliothèques de France, n.5, p.171-182.

GARDIN, J.C. (1974). Les analyses de discours. Neuchâtel: Delachaux et Niestlé.

GARDIN, J.-C. (1970). Procédures d'analyse sémantique dans les sciences humaines. In: POUILLON, J., MARANDA, P. Échanges et communications: mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème anniversaire. Paris, Mouton, p.628-657.

GARDIN, J.-C. (1979). Une archéologie théorique. Paris: Hachette, 1979.
GARDIN, J.C. (2001). Vers um remodelage des publications savantes: ses rapports avec les sciences de l'information. In: CHAUDIRON, S., FLUHR, C. Filtrage et résumé automatique de l'information sur les réseaux: actes du Chapitre Français de l'ISKO. Nanterre: Université de Paris X, 2001, p. 3-8.

GARDIN, J.-C. et al. (1981). La logique du plausible: essais d'épistémologie pratique em scieces humaines. Paris: Editions de la Maison des Sciences de l'Homme.

Jean-Claude Gardin e a análise documentária: trajetória da constituição de uma semiologia da representação

Maria de Fátima Gonçalves Moreira Tálamo
Universidade de São Paulo
mfgmtala@usp.br

Giovana Deliberali Maimone
Universidade de São Paulo
gdmaimone@usp.br

1 Introdução

Não me lembro quando, nem onde, só sei que foi há muito tempo. Em uma entrevista veiculada pela TV, Fellini, respondendo a uma pergunta sobre como Rimini, sua cidade natal, aparece nos seus filmes, afirmava que não conseguia considerá-la como fato objetivo, ela é antes uma dimensão da minha memória, quando estou lá sou agredido por fantasmas arquivados. De forma análoga falar da obra de Gardin, do seu impacto na consolidação dos estudos de análise documentária e na transformação da visão tradicional das operações no campo da informação, implica revisitá-la como memória afetiva, já que em torno dela uma coletividade de pesquisadores criaram laços fortes de formação, de pesquisa e de amizade. Não seria, no entanto, justo equacioná-la apenas nesses limites. Até para referendar o valor que assumiu na trajetória de pesquisadores da área de Ciência da Informação, resolvemos redimensioná-la, procurando revisitar o tema da análise documentária a partir de seus conceitos fundantes que lhe conferem um caráter reflexivo importante para apropriações outras que impactam nossa visão do fazer científico. Para isso resgatamos aspectos intervenientes na proposição do conceito que ainda estão instigando nossa percepção e motivando a curiosidade intelectual.

Isso posto, pretende-se resgatar aspectos intervenientes na proposição de análise documentária que a relaciona, de um lado, às características da ciência e de outro às características tipológicas da análise semântica. Dito de outra forma, a abordagem aqui feita sobre o referido conceito tratará das fronteira entre dois grandes

universos: um relacionado ao desenvolvimento das ciências humanas e sociais até os anos de 1970 e o segundo fortemente associado às alterações sobre o processo de interpretação documentária que no Brasil foi mote de estudos, análises e pesquisas frutíferas que acabaram por moldar a formação de parte dos profissionais atuais, que agora convido para revisitar alguns fragmentos do emblemático texto de Gardin "Procedures d'analyse sémantique dans les sciences humaines".

O conceito de análise documentária apresenta desde o significado de suas expressões - gestão e sistemas de informação, de um lado, e museus de outro, uma espécie de conflito entre o passado e o presente entre o novo e o velho e ao mesmo tempo possibilidade de superação deste conflito com a afirmação da possibilidade destes dois tempos e mundos convergirem em torno de uma ideia única relativamente conciliadora e oportuna mas nem sempre simples.

A trajetória do pensamento de Gardin que culmina com a proposição da análise documentária tem sua origem na observação do impacto de uso de métodos de análise semântica mais objetivos e rigorosos do que os instrumentos tradicionais que orientavam a interpretação empírica nas humanidades. De uma só vez, na década de 1970, Gardin reconhece a importância da linguagem na produção da ciência e a necessidade fundamental de um aparato metodológico interpretativo de natureza linguística. Com isso Gardin reconhece a linguagem como território do pensamento científico e não só de uma área disciplinar. Ela é comum a todas as áreas: para isso, cita a série de trabalhos aplicados em várias disciplinas que tratam de descrever a significação dos mais variados textos. Em suma, Gardin nos mostra que não prescindimos da linguagem para as operações investigativas e que tal, então, nela resgatar a função da linguagem?

Para que tal proposta não se afirme como observação ingênua, salienta-se que a linguagem não é apresentada como reflexo do pensamento, mero instrumento dele, ou ainda como meio de comunicação. Gardin chama a atenção para o fato de que a linguagem é representação. É como representação que se desenvolvem atividades de natureza metodológica fundamentais para o exercício investigativo. Nesse sentido, toda interpretação no universo científico participa ou compartilha da função representação. Dito de outro modo, toda interpretação, toda teoria relativa aos fatos do mundo, constituem uma semiologia. É justamente esse conceito que será explorado para o entendimento da noção de análise documentária.

2 Semiologia do fato científico

Na tradição saussuriana, a Semiologia define-se como a ciência dos signos. Este, por sua vez, é tradicionalmente apresentado, tautologicamente, como a relação entre um significante – imagem acústica – e um significado – imagem conceitual. – que denominamos significação. Embora a definição do termo seja difícil, reconhecem-se alguns consensos: o signo é sensível – o significante –, existe para uma comunidade determinada de usuários e o significado – que para o grupo, para o qual o significante torna-se sensível, assinala uma falta nela mesma. Diz-se, então, que o significante pode tornar-se sensível e o significante remete à parte ausente do signo. Já a relação entre significado e significante que denomina-se significação - não se confunde com a representação. De fato, para Ducrot e Todorov (1988), consideram que a representação está associada à abstração e esta, por sua vez, pode ser enunciada segundo diferentes gradações ou níveis. Infere-se que ao lado das propriedades já assinaladas, o sistema sígnico apresenta também propriedades representativas. Na linguagem verbal, o signo é dito, significa, mas, para além disso, também representa. A significação da palavra “chama” não a impede de representar “paixão/amor” na literatura e diferentes graus de abstração na Física associados às expressões fonte de calor e fonte de energia, respectivamente.

Observa-se, no exemplo acima, que a representação obtida por abstração está inserida no campo específico da ciência e não goza da mesma arbitrariedade da relação sígnica presente na linguagem comum. Isto significa que o pensamento científico opera com uma linguagem mais abstrata – relação signo/representação e, portanto, menos arbitrária como a existente entre o signo e a significação. Os estóicos explicavam tais associações através da distinção de três relações da parte perceptível do signo: o dizível, a coisa real e a imagem psíquica (DUCROT; TODOROV, 1988, p. 105).

Provavelmente esse aspecto ambíguo ou expandido do termo signo, é observado por Gardin, (1966, p. 171) ao defini-lo como símbolo

> [...] numa acepção deliberadamente geral, todo o símbolo que designa os elementos de um conjunto lexical (palavras-chave, descritores, termos de indexação etc.), sejam eles emprestados ou não da linguagem natural, “puros” ou “codificados”, alfabéticos ou numéricos etc.

Observa-se na transcrição acima o uso da palavra símbolo ao largo da definição estrita de signo, isto é, relação significante - significado. Isso decorre de recurso largamente usado pela linguística estrutural de diferenciar a definição estrita do signo de outras relações semelhantes mas diferentes que se originam na própria relação canônica. Este é o caso da representação que se associa mais ao símbolo do que ao signo. Embora sejam usadas de forma intercambiável, as duas denominações dão conta de diferentes dinâmicas de uso da palavra associadas não só aos processos de denotação e de conotação mas também e, principalmente, aos graus variados de abstração a que se submetem. Sendo assim, as relações entre significado e significante indicam união de elementos de naturezas diferentes – necessariamente a coisa e a forma: o signo é imotivado. Já o símbolo apresenta uma relação entre pares: normalmente tanto o simbolizante quando o simbolizado se apresentam como pares de uma relação não necessária. Aqui o universo não é o da significação mas o do sentido. Parece ser o simbólico o universo do pensamento científico, no qual as definições e a análise semântica cooperam para o desenvolvimento da capacidade de representação da ciência. Gardin (19170) não fala do signo no universo da ciência, ele afirma a existência de uma semiologia científica, orientando-se possivelmente pela semiologia ou semiótica que, essa sim, propõe o símbolo como unidade representacional do sistema.

Do mesmo modo em que esclarece a distinção entre significação e representação, a oposição arbitrário e motivado aplicada ao signo não prescinde, para a compreensão adequada da linguagem da ciência e da sua expressão em textos científicos, da distinção entre fato e fenômeno (GARDIN, 1970).

A expressão fato social refere-se ao objeto da Sociologia na sua origem e funciona como um recurso para o entendimento da distinção entre fato e fenômeno no interior da semiologia científica. Ele se define por um conjunto de características, é construído e não se reduz a nenhum fenômeno social. Ao construir seu objeto, a Sociologia indica um modo de formação do campo disciplinar das ciências humanas e sociais através da construção de um objeto de estudo, que denominaremos fato para harmonizar com a ideia genérica de pensamento científico. Granger (1970) examina as condições de existência de um fato, científico /semiótico, como produto de uma práxis (trabalho dinâmico) – e não de uma subjetividade passiva. Nessa perspectiva a razão desenvolve uma ação semiológica, isto é, uma manipulação de signos.

É nessa perspectiva que Gardin (1970), ao propor uma referência comum – o pensamento científico – para toda atividade que atribui significação ao mundo, acaba por associa-lo a uma semiologia. Assim, toda interpretação é também uma semiologia, o que nos faz integrar a análise semântica no aparato metodológico da prática científica, a partir de condições bem definidas da análise semântica, elencadas a seguir:

a) O objeto da análise, nesse caso, restringe-se aos documentos escritos, concebidos como textos gramaticais de uma língua natural. Semelhante objeto empírico pode ser submetido à interpretação empírica ou parafraseado – entenda-se representado – por uma metalinguagem autônoma;
b) O pensamento científico como sistema simbólico (semiologia) designa um conjunto de signos efetivamente construídos e enunciados e logicamente relacionados.

A forma de analisar os textos científicos – que Gardin denomina estudo comparado – recai sobre a totalidade do conjunto de signos. De modo específico, essa restrição relaciona-se à precisão das operações analíticas, que devem se apresentar explicitamente definidas para dar conta da passagem do texto original à sua representação.

Decorre dessas duas condições, que o sucesso da análise comparada, de natureza semântica, vincula-se a identificação dos pontos comuns e das diferenças entre as unidades do sistema, discriminadas nos sistemas simbólicos particulares. Como a diversidade do universo semiótico é um fato, o instrumento de análise deve situar os sistemas formais como aparato metodológico de análise/representação das línguas naturais e de seus produtos. Tal discriminação é necessária, uma vez que toda ciência é estrutural, isto é, torna a experiência objetiva, passível de ser apresentada como uma rede de elementos relacionados entre si. Com isso de algum modo garante-se que os elementos e as relações que mantem entre si constitua um sistema correlativo à experiência.

Por mais trivial que possa parecer, é justamente a diversidade de semiologias e a sua manifestação a partir da própria relação de significação que responde pela questão crucial da *análise semântica*: a palavra ou a unidade de significação não será nunca unívoca o que leva a definir o signo pela sua função de representação.

Neste sentido cabe ressaltar a definição atribuída à expressão *ordem semântica.* Para Gardin (1966, p. 175) a expressão compreende “uma ordem dos termos (ou seja, dos objetos ou noções que eles designam) que seja o reflexo de um conjunto de definições correntemente admitidas num grupo humano dado”. Isso reforça a ideia de semiologia científica e das suas unidades simbólicas.

Lembremos ainda que o processo de simbolização não se resume apenas a explicar a distinção entre significação e representação. Desde o estruturalismo saussuriano, distinguem-se duas formas de representação simbólica consideradas divergentes: a comunicação e a simbolização. Na perspectiva de Granger (1970), a comunicação de um conteúdo está comprometida com o vivido sinalizando para uma experiência concreta a qual faz referência – a parte perceptível do signo. Já a forma construída indica um conjunto de traços genéricos, obtidos por operações de abstração. Nessa perspectiva, ambos os processos são irredutíveis embora estejam intimamente associados: a relação entre os interlocutores de uma mensagem é comunicativa, a simbolização, por sua vez, designa a produção de objetos conceituais.

Portanto, o objeto científico não pode se apresentar pela face perceptível do fenômeno. Este, embora seja resultado de um recorte na realidade, é descritivo. Nele os seus elementos constitutivos encontram-se isolados e definidos. A definição semântica exemplifica esse aspecto já que apresenta os elementos definitórios da palavra por meio das propriedades internas e positivas. O máximo que se pode exigir da organização de um elenco de fenômenos é a sua classificação. Já um sistema formal, por sua vez, se caracteriza pelos aspectos pertinentes dos signos que o compõem delimitados de forma inequívoca e por uma construção simbólica realizada a partir de uma lista fechada.

Assim, enquanto as línguas naturais exercitam ambas as funções, comunicativa e de simbolização, de forma equilibrada, os sistemas formais apresentam um grau maior de abstração e uma comunicação mais limitada, vale dizer, restrita. Nos sistemas formais, a função comunicativa é secundária. Entende-se portanto o significado de expressões como “vulgarização da ciência” que sinaliza para o caráter “metalinguístico universal” das línguas naturais, fortemente associado a sua função de sistema modalizante primário e, portanto, de eficiente meio de comunicação.

De forma resumida, Gardin (1970) reconhece que o pensamento científico se desenvolve através de procedimentos de análise, dito também interpretativos.

Reconhece também, e não deixa de ironizar esse aspecto, que na maioria das vezes essas interpretações são empíricas. Interpretações empíricas estão associadas à totalidade da experiência, isto é, se relacionam ao vivido, ao perceptível e não dão conta do fato construído. Evidentemente, a semiologia do vivido é a base de toda prática inclusive a científica mas não é suficiente para fundamentar a ciência. A Semiologia a que se refere Gardin, exige a formalização do vivido, ou o reconhecimento de uma lógica própria, específica da prática científica que não se confunde com o vivido.

Assim, ao propor a prática científica como uma Semiologia, Gardin reconhece a necessidade de uma representação específica dos conteúdos dos textos escritos que não se confunde com a interpretação empírica que ocorre na maioria das vezes e que se submete à subjetividade da significação, dos fragmentos e da percepção da experiência. Lembrando ainda que a consistência da proposição da noção de representação como aparato do pensamento científico – base da semiologia – não privilegia a função comunicação, ou melhor, não a propõe como função primária do sistema sígnico.

3 Análise Semântica e Análise Documentária

Gardin (1970), como vimos, parte da noção de pensamento científico para apresentar como um dos seus recursos metodológicos a análise semântica. Como se disse, a dinâmica do pensamento científico, o seu universo, constitui uma semiologia. É preciso então esclarecer quais procedimentos estão aí inscritos.

Referindo-se ao pensamento científico de forma genérica, efetivamente Gardin não irá propor semiologias científicas associadas às diferentes áreas de especialidade. Embora não sinalize a discussão da interdisciplinaridade, seu compromisso é o de identificar a melhor opção metodológica como adjuvante do pensamento científico, ou ainda a operação linguística capaz de defini-lo. Reconhece para tanto uma tipologia de análises textuais: análise do conteúdo, análise estrutural (narrativa) e análise literária. Observa ainda que as tentativas de formalização recaem sobre as duas primeiras enquanto que a última – análise literária – encontrava-se ainda envolvida pela tradição da crítica textual.

Para contemplar uma compreensão mais ampla da função e procedimentos da análise no processo de construção da ciência, basta observar o que ocorreu na

Europa, especialmente na França, o espírito historicista e impressionista das análises dos textos das humanidades se vê comprometido, a partir da década de 1960, pelo estruturalismo na Etnologia e na Linguística. Lèvi-Strauss, Jakobson, Benveniste exploram em suas análises do fato cultural e literário suas estruturas narrativas e textuais que dão conta do seu funcionamento (TRASK, 2004). De um certo modo, as análises que são propostas dos fatos da cultura, entendidas na sua globalidade, acabam por desacreditar as interpretações empíricas. Como bem o observa Granger (1968, p. 138), no seu estudo sobre os mitos, a mitologia aí é o objeto de análise e não os mitos. Não se trata de uma reflexão sobre a visão do mundo do povo que criou um conjunto de mitos. Nesse caso prevaleceria a interpretação subjetiva a que se submete usualmente à apreciação por critérios de autoridade: a análise limita-se ao fenômeno e o fato não chega a ser constituído ou enunciado. A mitologia compreende portanto o estudo/análise do fato e não do fenômeno.

Evidentemente que não se trata apenas de validar ou apreciar a qualidade de uma interpretação ou de um comentário feito a partir de dados, textos ou quaisquer outros elementos da realidade. Trata-se sim de se reconhecer os argumentos sobre os quais eles se organizam. É interessante observar que a busca de um padrão para orientar a análise semântica está presente em diferentes momentos da investigação sobre a linguagem. Quase que invariavelmente, os movimentos nesse sentido, visavam o estudo da literatura, especialmente nas primeiras décadas do século XX, quando a Poética passa a expressar uma ambição científica. De fato, “seu objeto não é mais o conjunto das obras literárias, mas o discurso literário enquanto princípio de engendramento de uma infinidade de textos” (DUCROT; TODOROV, 1988, p. 84).

Observa-se portanto que as análises ou interpretações empíricas tem um papel importante no desenvolvimento da Poética mas não constituem seu núcleo. De maneira semelhante, a noção de texto – definido pela sua autonomia e fechamento – acabará por provocar diferentes análises – narrativa, temática e retórica – cada uma delas fundada em uma abordagem específica do texto. No caso da narrativa, por exemplo, as análises dos contos populares realizadas por Propp tiveram importância fundamental para a sua definição como o “texto referencial com temporalidade representada” (DUCROT; TODOROV, 1988, p. 268) que passa a ser o conceito que orienta a análise propriamente dita, que versa sobre noções de transformação, mediação, ação, tipologias de personagens e de sequenciais narrativas, etc.

Embora, esse contexto, torne mais evidente a dualidade entre interpretação e análise propostas por Gardin, a explicação mais definitiva pode ser obtida a partir de Granger. Em 1968, na sua obra “Essai d’une philosophie du style”, Granger desenvolve uma discussão sobre o processo de estruturação e acaba por definir o estilo como uma solução individual provocada pelas dificuldades encontradas por todo o trabalho de estruturação. Isso porque a passagem do amorfo para o estruturado não se realiza por imposição de uma forma exterior: a prática científica encontra a resistência da experiência – do vivido – e a estruturação da experiência só pode ser desenvolvida por um trabalho – a própria prática científica – que relaciona forma e conteúdo. Explica ainda como se desenvolve o trabalho científico na matemática e na linguagem e como o indivíduo se comporta frente à solução estruturante: produzindo um estilo e não uma solução subjetiva associada diretamente ao vivido. Para dar conta da distinção entre o vivido e a estruturação – duas formas significantes diferentes, propõe duas formas de manipulação sígnica, duas semiologias, que denominaremos semiologia do vivido (da experiência) e semiologia científica ou das formas estruturantes. Na terminologia de Granger a semiologia do vivido é produtora do sentido e a científica produtora da significação. Essa oposição corresponde ao que Gardin denomina, respectivamente, interpretação empírica e representação. A análise documentária é uma forma de manipulação sígnica, de natureza representacional que constitui a semiologia científica. Essa semiologia, portanto, é informacional.

Uma semiologia do fato ou da constituição objetiva de sistema significante é para Granger (1966) de natureza informacional, à medida que trata informacionalmente a realidade. A distinção que aqui se estabelece é entre o que é estruturável (*stricto-sensu*) e os sistemas significantes. De fato, é impróprio admitir que esses dois modelos possam se desenvolver no interior da semiologia da ciência.

Considerando-se o que foi dito, entende-se que as “análises interpretativas” das estruturas significativas não podem ser confundidas com a análise da estrutura simbólica (informacional, do fato). Efetivamente, as análises feitas sobre o vivido (visão de mundo) ou análises significativas não se deixam levar pelo conceito de estrutura.

Seguindo esse ponto de vista, Gardin (1970) propõe a análise documentária como uma manifestação da análise semântica, do conteúdo dos textos científicos “à la manière des documentalistes” (GARDIN, 1970, p. 630). Para essa proposta baseia-se em algumas técnicas de análise automática do conteúdo ou estrutura na

antropologia. Embora tradicionalmente os métodos de análise documentária reduzam-se às operações empíricas de atribuição de “palavras –chave” aos textos, livros, etc., fundadas mais na experiência e hábito do que ao procedimento explícito (GARDIN, 1970, p. 631). Com a automação, a análise automática passa a identificar a formalização da interpretação semântica. Assim, os textos científicos passam a ser substituídos por paráfrases ou por representações elaboradas por meio de uma metalinguagem autônoma, que define os elementos constitutivos da paráfrase.

Portanto, a análise documentária, através do seu sistema objetivável estabelece, uma correspondência entre os fenômenos, ou o vivido, e as representações da metalinguagem. A análise documentaria é então parte fundamental da Semiologia informacional, a qual no seu desenvolvimento e transformação integrará o conceito de conhecimento associado à sua produção.

4 Conclusão

A língua perfeita, qualquer que seja o entendimento que se tenha dela, faz parte do imaginário coletivo e, não raro é um tema perseguido, provavelmente porque a narrativa babélica nos faz crer na diversidade negativa, limitante e inoperante. Como bem observa Eco (1993, p. 406-407) discorrendo sobre as consequências da diversidade das línguas, a qualifica com um fenômeno extremamente positivo “que permitiu a fixação dos assentamentos, o nascimento das nações e sentimento da identidade nacional. A diversidade não é incontornável e pretender universalizá-la não deixa de ser um processo ao mesmo tempo ambicioso e inócuo. Do mesmo modo, submeter um texto científico à interpretação empírica é desqualificá-lo enquanto tal e, empobrecê-lo informacionalmente.

A análise semântica objetivável reconhece padrões subjacentes, semelhantes, de pensamento. Lévi-Strauss (2003), um dos mais importantes intelectuais do século XX, ao instituir a antropologia moderna afirma a necessidade da busca consistente de padrões subjacentes de pensamento em todas as formas de atividade humana. Para ele, o homem organiza o conhecimento em pares binários e opostos que se estruturam segundo a lógica própria de cada universo, campo, sociedade, etc. Essa análise lógica, evidencia, por exemplo, que o texto da ciência fundamenta-se nas definições dos conceitos utilizados. Tal restrição de forma, de natureza lógica, tende a facilitar a formatação, o armazenamento e especialmente o uso do conteúdo dos textos em um

mundo onde o pesquisador está cada vez mais oprimido pelo volume de dados. A representação, no sentido que lhe confere Gardin, funciona com um controle da significação do vivido.

Dito de outro modo, o cientista constrói um texto que responde às necessidades de transparência dos argumentos, contemplando soluções para as mais diferentes questões. O discurso do autor depende sem dúvida das formas de organização da cultura mas a sua estrutura científica deve ser independente do discurso do autor, integra o discurso investigativo que não é, no entanto, independente do autor. Sob essa ótica, a Semiologia informacional que desenvolve a análise lógica distingue duas operações realizadas sobre o texto: a interpretação, de natureza autoral, e a tradução, de natureza científica.

Enfatiza-se finalmente que a representação, tal como definida nesse texto não mantém um elo mecânico com a função cognitiva. Não raro a linguagem, dada sua complexidade, tem suas funções exploradas segundo o objetivo do sistema construído. O controle da significação promovido pela terminologia visa primordialmente controlar a dispersão e facilitar a comunicação entre os especialistas, é vital que se controle a linguagem: Cabré (1993, p. 43, tradução nossa) argumenta que “a terminologia é uma ferramenta básica para a comunicação especializada (uma boa terminologia dá garantias a seus usuários sobre a precisão e eficácia da comunicação)”. Nesse sentido, a terminologia pode auxiliar nos seguintes processos: permutar informação entre diferentes grupos; fornecer uma descrição adequada sobre um domínio; melhorar o entendimento de um domínio; evitar distorções cognitivas; permitir reuso de conceitos em diferentes domínios.

Nesse momento, integra-se às nossas memórias o contexto fundante da análise documentária que nos permitirá, sem dúvida entender a trajetória que Gardin percorre em suas investigações, especialmente a partir da década de 1980.

Referências

CABRÉ, M. T. **La terminología**: teoría, metodología, aplicaciones. Barcelona: Antártida/Empúries, 1993. 529p.

DUCROT, Oswald; TODOROV, Tzvetan. **Dicionário Enciclopédico das Ciências da Linguaguem.** São Paulo: Editora Perspectiva, 1988.

ECO, Umberto. **A busca da língua perfeita**. Bauru: EDUSC, 2001.

GARDIN, Jean-Claude. Elements dun modele pour la description des leqxiques documentaires. **Bulletin des Bibliothèques de France**, v. 11, n. 5, p. 171 – 182, 1966.

GARDIN, Jean-Claude. Procedures d'analyse sémantique dans les sciences humaines. In: POUILLON, Jean; MARANDA, Pierre (coord). **Échanges et Communications.** Paris: Mouton, 1970.

GARDIN, Jean-Claude. Document analysis and linguistic theory. **Journal of Documentation**, v. 29, n. 2, p. 137 – 168, 1973.

GARDIN, Jean-Claude. Lecturas plurales y ciencias singulares de la literatura. **Diógenes**, Universidad Nacional Autónoma de México, 1982, v. 118, p. 5 – 16. Verano/1982.

GRANGER, Gilles-Gaston. **Essai d'une philosophie du style.** Paris: Librairie Armand Colin, 1968.

LÉVI-STRAUSS, Claude. **Antropologia Estrutural**. Rio de Janeiro: Ed.Tempo Brasileiro. 2003.

TRASK, R. L. **Dicionário de linguagem e linguística**. São Paulo: Contexto, 2004.

Linguagem documentária na ótica de J.-C. Gardin

Marilda Lopes Ginez de Lara
Universidade de São Paulo
larama@usp.br

1 Introdução

Discursos recentes da Ciência da Computação falam em ‘acrescentar semântica’ à recuperação. Essa não é, entretanto, uma preocupação recente: Jean-Claude Gardin, há cerca de 50 anos atrás, buscava resolver problemas de um sistema ou codificação cujos termos necessariamente deveriam corresponder a uma definição suficientemente detalhada para que não subsistissem dúvidas quanto à sua respectiva significação (GARDIN, 1968). A esse sistema Gardin denominou ‘metalinguagem’, qualificando assim a linguagem usada para tratar e recuperar textos técnico-científicos.

O enfrentamento das questões da significação continua sendo um desafio, e revisitar o trabalho de Gardin pode não só permitir conhecer as bases teórico-metodológicas da representação do conhecimento por ele propostas ou sistematizadas, como realizar sua avaliação frente às iniciativas atuais. Necessário também ressaltar a importância de seu legado para o ensino: a produção de Gardin faz diferença na formação acadêmica e profissional em Ciência da Informação, porque assegura o entendimento dos fundamentos de uma atividade não raras vezes visto como meramente instrumental.

Em muitos aspectos, as reflexões de Gardin estavam à frente de sua época demonstrando, além do profundo conhecimento da literatura sobre a linguística e a lógica, pioneirismo na discussão e experimentação na área da semântica. Para a pesquisa de hoje, Gardin é importante porque toca em problemas ainda não

totalmente resolvidos e porque situa convenientemente a problemática da significação no universo da representação e da interpretação em linguagem.

Para organizar a exposição, após esta introdução localizaremos o termo 'linguagem documentária' frente a outra terminologia e discutiremos sua classificação como metalinguagem para destacar a ênfase que Gardin conferia à relação metalinguagem-*corpus* discursivo. Apresentamos, depois, uma tipologia dos léxicos, onde Gardin compara as ferramentas documentárias aos léxicos de apoio e, em seguida, os principais componentes da linguagem documentária, caracterizada por Gardin como metalinguagem. Procuraremos mostrar aspectos da discussão do autor sobre a análise semântica, destacando algumas de suas observações sobre a pertinência das propostas da linguística à época. Nas considerações finais destacaremos a atualidade de suas preocupações, apontando para a proximidade de algumas de suas propostas com as formuladas recentemente.

2 Linguagem documentária: denominação e classificação

O termo linguagem documentária divide seu espaço com outros termos de significado aproximado – linguagem de indexação, vocabulário controlado etc. – cada um deles priorizando um aspecto diferente. Mais recentemente, a literatura em inglês registra o termo *Knowledge Organization Systems – KOS,* para agrupar uma série de recursos e instrumentos com funções diversas, mas também próximas: textos não estruturados (do tipo *'abstract'*), listas de termos e conceitos (folksonomias, listas de autoridade, listas de *tags*, dicionários, glossários), listas estruturadas de conceitos (vocabulários controlados, sistemas de classificação, listas de cabeçalho de assunto, taxonomias, tesauros, ontologias) e, enfim, estruturas de relacionamento (mapas conceituais e modelos de referência de formas de organização) (SOUZA; TUDHOPE; ALMEIDA, 2012).

O W3C opta pelo termo genérico vocabulário, afirmando que na Web Semântica, os vocabulários definem conceitos e relacionamentos ou termos usados para descrever e representar uma área de interesse. Moreiro González (2011) também defende o uso da expressão vocabulários semânticos, justificando sua escolha pela atualização da terminologia no contexto da Web.

As variações devidas a ênfases diferentes, a ampliação da abrangência via *KOS,* ou a atualização específica da expressão linguagem documentária nesse âmbito, não apaga a diferença básica que marca o uso do termo 'linguagem' frente às outras denominações: a referência à linguagem natural, sistema complexo de comunicação cujas unidades têm seu valor definido relacionalmente no interior do sistema, ou seja, significam em função do conjunto (BENVENISTE, 1991). Vocabulários não são necessariamente estruturados: mais frequentemente são listas de ocorrências de um determinado *corpus* discursivo (GREIMAS; COURTÉS, 2008); já as linguagens documentárias o são, e sua melhor caracterização como tal depende da presença e da sofisticação de suas relações semânticas, bem como da potencialidade em gerar sintagmas via relações sintáticas. Não é sem razão que uma das definições mais precisas de linguagem documentária é a que a define a partir de seus elementos constituintes: um léxico, uma rede paradigmática e uma rede sintagmática (CROS; GARDIN; LEVY, 1968), elementos esses que respondem pela estrutura da ferramenta. O vocábulo 'documentária' (ou documental, em algumas versões), restringe, de fato, o termo linguagem documentária a um universo de documentos, podendo-se discutir sua flexibilização.

Na busca de um modelo de Análise Documentária, Gardin se refere a três fatores: "primeiro, à definição de análise documentária como extração de significado de um texto, o que implica uma referência a algo que é 'tirado' do texto e designado por símbolos que não são necessariamente os encontrados no texto – a saber, conceitos e suas interrelações, designados por símbolos *ad hoc* (descritores, indicadores de função, etc.). Para o sistema de símbolos usados para expressar o conteúdo, [o autor] propôs o nome de 'metalinguagem' " (GARDIN, 1969, citado por GARDIN, 1973, p.144). Em segundo lugar, à existência de similaridade entre as metalinguagens existentes à época, atribuindo suas variações às diferenças entre os campos e aos objetivos da representação. Gardin identifica, porém, analogias relativas: a) aos procedimentos seguidos para definir a unidade de análise, b) ao estabelecimento de relações analíticas entre os termos do conjunto, de modo a construir a organização semântica do léxico em questão, e c) à provisão de relações sintáticas entre os termos para conferir estrutura lógica à representação metalínguística das unidades sob análise. Em terceiro lugar, o autor se refere à distinção entre relação analítica (ou semântica) *versus* sintática (ou lógica). As relações semânticas são mais estáveis – como as que ocorrem em 'inseticidas',

diferentemente das relações ‘imprevisíveis ou sintáticas’, que são circunstanciais, acidentais, como em Drogas (Instrumento), Insetos (Paciente) (GARDIN, 1973). No entanto, a diferença entre as duas só tem fins práticos: “sob o plano sintático são veiculadas as relações novas, que são objeto do discurso, portanto não estabilizadas. Passa-se, no campo da ciência, progressivamente do plano sintático para o plano semântico: as relações novas, nesse sentido, ganham estabilidade e passam a constituir as relações semânticas“ (GARDIN, 1973, p.144-145).

Tais fatores, em conjunto, justificam, para o autor, caracterizar a linguagem documentária como metalinguagem, porque ela põe em relevo o processo de construção e de transformação de textos que guardam contiguidade semântica entre si. Essa metalinguagem tem função normalizadora, operando semanticamente no tratamento de homonímias, homotaxias, homografias, polissemias etc. (GARDIN, 1970). Embora Gardin considere perigoso emprestar uma palavra que tem um sentido distinto na lógica, o termo metalinguagem é interessante, porque remete simultaneamente à linguagem de partida e à sua representação (em linguagem). Esses fatores justificam, também, sua posição relativamente à linguística chomskyana quanto à unidade de análise e quanto às suas contribuições para prover a linguagem com um esquema de articulações.

A necessidade de uma metalinguagem, para o autor, se deve às ‘anomalias’ da linguagem natural do ponto de vista semântico: sinonímia, homonímia, homografia, polissemias; alotaxias (variações de formulação das frases), equivalências mais complexas (entre a designação e sua definição, por exemplo). Os sistemas simbólicos estranhos à linguagem natural apresentam irrregularidade de correspondências entre significantes e significados fundadas na invariância relativa dos significados nos domínios considerados (GARDIN, 1970, p. 632). “A metalinguagem … não é outra coisa que esse sistema de símbolos, também chamado linguagem documentária, linguagem de recuperação especializada *(‘information retrieval language’*, em inglês)” (GARDIN, 1970, p.634).

3 Oposição entre os léxicos: lexicografia natural e lexicografia documentária

Não menos importante para as reflexões sobre a organização da informação e do conhecimento é a proposta que posiciona as linguagens documentárias no conjunto dos ‘léxicos’. Isso demonstra que elas não foram vistas por Gardin como produtos totalmente separados da linguagem natural. A tipologia, anterior à proposta

de metalinguagem, já sinaliza a ênfase que o autor dará, no futuro, à relação entre linguagem natural e linguagem documentária e a especificidade das ferramentas de representação.

Os 'léxicos documentários' compreendem "todo conjunto de signos (palavras naturais, símbolos alfa-numéricos etc.) organizados ou não, que servem para construir representações indexadas de certos documentos" (GARDIN, 1966). Os produtos da 'lexicografia documentária' e os da 'lexicografia natural' (glossários e tesauros de língua), auxiliares da atividade documentária, são hoje reconhecidos como KOS.

Na construção das árvores que os organizam, a cada oposição, são discriminadas as diferenças entre os léxicos considerados. Os traços distintivos que permitem descrever as dicotomias são a correspondência com a linguagem natural (explícita ou implícita), a organização ou não dos léxicos (listas e classificações), a dimensão considerada (unidimensionais e pluridimensionais, unidimensionais reais e aparentes), o tipo de organização privilegiada (semântica, sintática ou mista). Sob mistas, as classificações são organizadas pela forma de organização (quase-hierarárquicas, analítico-sintéticas e facetadas), sendo que as duas primeiras ainda são organizadas segundo a frequência de um termo na ferramenta (unívocas, multívocas).

As tipologias são sempre interessantes como meio para permitir visualizações, mas o mais importante dessa proposta de Gardin é que ela, ao recorrer a dicotomias num crescente processo de descrição semântica, reafirma o caráter estrutural da classificação e antecipa sua natureza metalingüística.

4 Componentes da linguagem documentária

Focalizando as linguagens documentárias em detalhe, Gardin identifica, no léxico, o componente mínimo da metalinguagem, conjunto de termos extraídos de uma dada linguagem natural (palavras-chave ou descritores). Um segundo componente é a estrutura paradigmática, conjunto de dados relacionados entre si *a priori* no léxico (GARDIN, 1965, citado por GARDIN, 1973). Esses dois componentes caracterizaram a maior parte das linguagens documentárias na sua configuração mais tradicional. Um terceiro elemento, uma estrutura sintagmática, foi adicionada por Gardin e equipe (CROS; GARDIN; LÉVY, 1968) na elaboração do SYNTOL – *Syntagmatic Organization Language*, para conferir maior flexibilidade e potencialidade

à ferramenta, bem como para dar conta da imensa diversidade de dados relacionados observada no processo de análise dos documentos (GARDIN, 1973).

O autor relativizou, no entanto, a distinção entre as duas estruturas (paradigmática e sintagmática) afirmando que ela era desejável para prover 'pontes' entre uma metalinguagem e outra quando da conversão intra-metalinguagem ou inter-metalinguagens (Aqui, uma antecipação da preocupação com a compatibilidade entre linguagens). A unidade relacional básica foi então definida, com base nas duas estruturas, como o sintagma mínimo ‘Ri (x,y)’, algoritmo de análise para formalização, onde x e y são termos do léxico e R é uma relação binária cujo significado é especificado por i. Regras de orientação são adicionadas para evitar ambiguidades.

Por meio desse sintagma mínimo é possível representar tanto relações paradigmáticas como sintagmáticas: no primeiro caso, relações estáveis entre os termos; no segundo, relações ‘reais’ que pressupõem interdependência entre dois significados. Tais relações são classificadas como dinâmicas (relações consecutivas ou de causalidade), estáticas (superposição associativa e predicativa entre conceitos), ou formais (coordenativas, comparativas e aproximativas).

O SYNTOL permite formular frases ao ligar duplas de palavras com base em suas relações mútuas de dependência. O sistema prevê a indicação do tipo de relação lógica existente entre as palavras consideradas (relação predicativa, associativa, consecutiva ou coordenativa), a sinalização do sentido da leitura do sintagma e a determinação das categorias formais respectivas (determinação da categoria num nível temático – lugar, tempo, modo – mais amplo do que o nível dos sintagmas, de modo a deixar claro o contexto). Operadores sintáticos são usados para evitar ambiguidade de interpretação, quando necessário, ou para ligar sintagmas próximos (CROS; GARDIN; LÉVY, 1968).

A proposta partiu, entre outras observações, da verificação de que as relações binárias são frequentemente utilizadas pelas metalinguagens para realizar a organização semântica, do mesmo modo que os indicadores de função (*role indicators*) o são para expressar relações sintáticas: os marcadores funcionais (predicados unários), cada um deles implicitamente referido a um ou vários marcadores semânticos numa dada representação metalinguística. Segundo Gardin, a vantagem do formato binário é que ele pode orientar cálculos lógicos para realizar deduções e inferências no processo de busca, quando especificações lexicais ou sintáticas devem ser gradualmente ampliadas para aumentar a revocação na

recuperação, ou ainda, na conversão de uma metalinguagem para outra, quando os mapeamentos (árvores sintagmáticas ou gráficos SYNTOL) das duas implicarem árvores de relações diferentes (GARDIN, 1973).

A experiência do SYNTOL, aqui apresentada de forma muito breve, mostra o avanço das pesquisas em Organização e Representação do Conhecimento numa época em que a maior parte das operações de sintagmatização era, no máximo, resolvida a partir do uso da lógica booleana.

5 A análise semântica

Ao procurar por procedimentos destinados a colocar em evidência a significação de textos mais variados, Gardin (1970) mostra, já na década de 60 do século XX, a importância da análise semântica, quando os métodos de análise documentária eram reduzidos a operações empíricas de atribuição de palavras-chave a um texto. O autor discute os limites da linguística chomskyana – a teoria gerativo-transformacional – e se queixa de que raramente os linguistas falam de análise semântica.

Gardin critica, na teoria linguística dominante à época, a prioridade dada à análise sintática e à 'estrutura de superfície', que exerceriam papel prioritário na determinação do significado. Fala dos limites da estrutura profunda, questiona a sentença como unidade padrão de análise, o uso de categorias gramaticais (nomes, advérbios, adjetivos etc.) e de funções gramaticais (sujeito, objeto etc.) como meios para a expressar o significado. Cita outros autores (cujas contribuições foram publicadas entre 1968 e 1972) que também questionaram as propostas chomkyanas, a exemplo de Wilks e Hutchins.

Por outro lado, Gardin refere-se a uma série de experiências levadas a efeito por linguistas no mesmo período (início da década de 1970): a proposta de estruturas pré-lexicais (Postal), a adoção de 'estratégias semânticas' para guiar a interpretação de sentenças (Bever), a definição de 'estruturas formacionais' antes da análise dos enunciados (Grimes), uma unidade de análise menor do que a frase (Wilks) e, ao contrário, unidades mais amplas para prover uma melhor interpretação (sugeridas individualmente por Langendoen, Lakoff, e Thompson). O autor chama atenção para as pesquisas desenvolvidas na então URSS (Mel'chuck e Zholkovski, entre outros), que propuseram modelos de linguagem sofisticados com base em estruturas

semânticas e que, diferentemente de Chomsky que trabalhava do texto para o significado, iam em direção contrária, das estruturas conceituais para as expressões linguísticas (GARDIN, 1973).

Gardin se reporta também a outras experiências que considera interessantes, como a análise do discurso, a valorização da 'proposição lógica', os conceitos lógicos como conceitos que subjazem à fala, as redes semânticas (*semantic networks)* que incorporam informação relacional ou combinatorial 'sobre o mundo' (Lakoff; McCawley; Kuroda). Do mesmo modo, refere-se às 'pressuposições' da linguagem, ou regras de seleção que orientam o comportamento linguístico (citando Bever, Fillmore e Langendoen), as quais o autor situa em paralelo à noção de organização semântica ou paradigmática usada na análise documentária. Acrescente-se também a referência às categorias usadas na formulação de preposições (originalmente da gramática de casos, proposta por Fillmore, em 1968) reformuladas por Langendoen (com as *'roles structures'*), por Bever ('*basis relational structures')*, e por Wilks *('mental templates'*).

Toda a crítica feita à linguística gerativa transformacional foi feita em favor da lógica, não a lógica filosófica, mas a lógica simbólica, técnica. Ela forneceria o meio adequado de expressar a organização do pensamento, para categorizar os vocabulários (Agente, Paciente, Instrumento, Objetivo etc). Sobre as categorias, Gardin ressalta a prática documentária como antecipadora de muitas das propostas de uma semântica aplicada.

6 Considerações finais

Como foi possível verificar, são várias as contribuições de Gardin para a compreensão e desenvolvimento das linguagens documentárias e para a representação de textos técnico-científicos. Seu trabalho, no entanto, não é isento de críticas: ao mesmo tempo em que ele era um homem à frente de seu tempo – enxergando na semântica os problemas para a veiculação da significação, enfatizando a importância da estruturação das metalinguagens etc., - devemos lembrar que ele foi inflexível quanto à abordagem dos discursos das ciências humanas de um modo geral, criticando-os por sua falta de objetividade e clareza a partir da referência da organização própria dos discursos das ciências naturais. Gardin enxergava como 'anomalias' a expressão da riqueza e complexidade do vocabulário

do discurso das ciências humanas, procurando meios para alcançar formas que retirassem o excesso de retórica que os caracterizava.

Inegável, no entanto, a contribuição do autor para o aprimoramento dos procedimentos da atividade com documentos, como se fosse necessário experimentar uma fase de rigor para chegar a revê-la, depois (como a afirmação de Wittgenstein, no *Tractatus*: “Deve, por assim dizer, jogar fora a escada depois de ter subido por ela” (WITTGENSTEIN, citado por GRAYLING, 1966, p.65).

Para finalizar, ressalto aqui a atualidade do projeto de Gardin quando se observa a similaridade de princípios que subjazem à proposta do sintagma ‘Ri (x,y)’, feita na década de 60 do século XX, e o RDF – *Resource Description Language*, linguagem para representação de metadados no formato de sentenças que relacionam entidades e propriedades. Não menos interessantes são suas preocupações sobre a compatibilidade e o mapeamento de metalinguagens, que correspondem hoje às iniciativas para promover a interoperabilidade entre vocabulários.

Ficaram fora desta exposição muitas contribuições não menos importantes e relacionadas, que seguiram as reflexões sobre as metalinguagens e as experiências sobre automação das análises. Uma delas é a proposta de esquematização de raciocínios ou ligações lógicas entre proposições descritivas (fatos e argumentação) e proposições interpretativas. Muito resumidamente, trata-se de um sistema que, a partir de uma base de conhecimentos, vai dos dados à interpretação através de um aparato lógico de cálculo, um motor de inferências, como um sistema especialista de inteligência artificial (GARDIN, 1987). Produto da análise logicista, é um dispositivo de maior formalização e sofisticação que permite passar, de uma indexação mais simples, à representação e à inferência (e interpretação). Mas esse é outro assunto.

Referências

BENVENISTE, É.. *Problemas de linguística geral, I.* Campinas: Pontes; Ed. UNICAMP, 1991.

CROS, R.-C.; GARDIN, J.-C.; LÉVY, F. *L’automatisation des recherches documentaires*: un modele général “Le SYNTOL”. 2. ed. revue et augmentée. Paris: Gauthier-Villars, 1968.

GARDIN, J.-C. Elements d’un modèle pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France,* 1966, p.171-182.

GARDIN, J.-C. Procédures d'analyse sémantique dans les sciences humaines. In: POUILLON, J. et MARANDA, P., orgs. *Échanges et communications:* mélanges offerts à Claude Lévi-Strauss à l'occasion de son 60ème. anniversaire (Tirage à part). Paris: Mouton, 1970, p.628-657.

GARDIN, J.-C. Document analysis and linguistic theory. *The Journal of Documentation*, v. 29, n.2, p.137-168, June 1973.

GARDIN, J.-C. Éléments d'un modèle pour la description des lexiques documentaires. *Bulletin des Bibliothèques de France*, v.11, n.5, p.171-182, 1966. Disponível em: <http://bbf.enssib.fr/consulter/bbf-1966-05-0171-001> Acesso em jun.2015.

GARDIN, J.-C. *Systèmes experts e publications savantes*. London: The British Library, 1987.

GRAYLING, A.C. *Wittgenstein*. São Paulo: Ed. Loyola, 1966.

GREIMAS, A.J. & COURTÉS, J. *Dicionário de semiótica*. São Paulo: Contexto, 2008.
MOREIRO GONZÁLEZ, J.A. *Linguagens documentárias e vocabulários semânticos para a web*: elementos conceituais. Salvador: EDUFBA, 2011. 128p.

SOUZA, R.R.; TUDHOPE, D. ; ALMEIDA, M.B. Towards a taxonomy of KOS: dimensions for classifying Knowledge Organization Systems. *Knowledge Organization*, v.39, n.3, p.179-192, 2012.

Da análise documentária à terminologia: percurso teórico e metodológico

Vânia Mara Alves Lima
Universidade de São Paulo
vamal@usp.br

O objetivo desta comunicação é apresentar o percurso teórico e metodológico percorrido em nossa pesquisa sobre a utilização da terminologia de um domínio do conhecimento como referente para a construção e gerenciamento de linguagens documentárias, mais especificamente os vocabulários controlados e os tesauros.

Esse problema de pesquisa começa a se delinear na década de 90 durante nossa prática profissional no processamento técnico de um dos maiores acervos bibliográficos da Universidade de São Paulo, o Serviço de Biblioteca e Documentação da Faculdade de Filosofia, Letras e Ciências Humanas, que nos apresentava diariamente questões sobre a organização, representação e recuperação de informações produzidas sobre um mesmo assunto por diferentes domínios do conhecimento, e que consequentemente abordavam diferentes pontos de vista.

Quando nos transferimos para o Serviço de Biblioteca e Informação da Faculdade de Arquitetura e Urbanismo da USP, nos deparamos com outro problema relacionado à representação da informação, ali o conhecimento produzido em um mesmo domínio era representado por diferentes designações em função, muitas vezes, do suporte em que se encontrava registrado, isto é, livros, mapas, projetos de arquitetura, slides, fotos, artigos de revista.

A busca de fundamentação teórica para a solução de problemas práticos nos levou aos trabalhos de Smit, Tálamo, Cunha, Lara, Kobashi, Guimarães e Fujita, fundadores do Grupo Temma que tem na obra de Jean Claude Gardin seus alicerces teóricos, mas que ampliaram consideravelmente os horizontes da denominada análise

documentária, seus processos e seus produtos, principalmente no que se refere à construção das linguagens documentárias.

Os estudos do Grupo Temma demonstraram que a Análise Documentária definida por Gardin (1981) como “um conjunto de procedimentos para expressar o conteúdo dos documentos científicos sob formas destinas a facilitar a consulta” é uma disciplina metodológica que sugere procedimentos para a análise de textos com o objetivo de selecionar conteúdos informativos os quais poderão ser representados, recuperados e disseminados (TÁLAMO, LARA, KOBASHI, LIMA, 1992 ; LARA, 2011) independente do suporte em que são registrados. Ao conceito de linguagem documentária como “um conjunto de termos utilizados para representar certo conteúdo de documentos científicos com fins de classificação ou de pesquisa retrospectiva de informações (CROS, GARDIN, LEVY 1968)”, o Grupo Temma acrescenta outras características que enfatizam seu caráter de “linguagem”, como uma estrutura, cujos termos devem necessariamente estar relacionados para que possam significar de modo determinado, funcionando como veículo de comunicação que represente os campos conceituais respeitando a cultura da comunidade a qual serve (VOGEL, 2004).

Neste quadro é que se insere nossa abordagem linguística e terminológica tanto sobre a elaboração da informação documentária, entendida como o resultado da representação do conteúdo de um documento, quanto da utilização das terminologias dos domínios como referente para a construção de linguagens documentárias.

No que se refere à informação documentária, procuramos esquematizá-la a partir das aproximações com os referenciais linguísticos, identificando-a como um signo documentário, analogamente ao signo linguístico, definido por Peirce (1977) como aquilo que, sob certo aspecto ou modo, representa algo para alguém e está no lugar de alguma outra coisa sob algum aspecto ou capacidade, ou seja, a informação documentária se propõe no lugar do conhecimento registrado e, portanto da mesma maneira que o signo linguístico é passível do processo de semiose, denominada por Lara (2006) de semiose documentária..

Entendendo-se a semiose como a construção do significado pelo intérprete, no interior de um contexto determinado, pode-se inferir que a semiose documentária é a construção de significados fundamentados em referenciais terminológicos os quais remetem às estruturas conceituais dos domínios. Portanto,as referências contextuais

de produção da informação documentária são essenciais para que ela efetivamente represente um conjunto de enunciados verdadeiros sobre o conhecimento registrado, os quais são compilados nas definições de cada termo presente na terminologia de um domínio.

Devemos esclarecer que quando falamos em terminologias estamos nos referindo à terminologia enquanto produto, isto é, ao conjunto de termos de uma especialidade, o qual é elaborado, através da utilização de normas terminológicas, propostas pela Terminologia enquanto disciplina que se ocupa dos termos especializados (CABRÉ, 1995).

Enquanto a terminologia do domínio é o reflexo formal da organização conceitual de uma especialidade e meio inevitável de expressão e comunicação profissional, o qual garante a transmissão do conhecimento, a linguagem documentária tem por função normalizar a busca e garantir a recuperação desse conhecimento registrado através da elaboração da informação documentária. Assim, entende-se que esses dois instrumentos são complementares.

Segundo Gardin (1981), informação documentária é produto da análise documentária, resultado de uma operação semântica formulada dentro de uma linguagem documentária, que transforma um texto original em uma ou mais palavras chaves, e que mesmo apresentando, na linguagem documentária, a mesma forma que na linguagem natural, não tem necessariamente todas as acepções presentes em um dicionário geral da língua

Nesse contexto inferimos a informação documentária enquanto uma representação do conteúdo do documento, a partir dos conceitos do domínio a que pertence, os quais designados pelos termos desse domínio servem como referente para os descritores da linguagem documentária, pois contextualizam a sua significação a partir de uma práxis nesse domínio (LIMA, 1998).

Na figura a seguir esquematizamos a informação documentária representada pela tríade conceito/termo/descritor resultado de processo de semiose documentária aliada à práxis em um domínio do conhecimento.

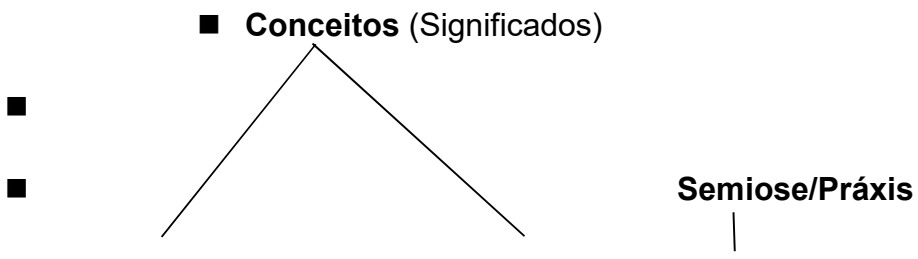


Figura 1 – A informação documentária

Esse esquema da informação documentária atende à colocação de Smit (2000, p.28) de que para que a informação documentária funcione efetivamente e corretamente como mediadora, entre o usuário e o estoque de informação, o usuário deve ter condições para contextualizá-la e decodificá-la, pois ao utilizar a terminologia do domínio como referente para a linguagem documentária, possibilita ao usuário o reconhecimento dos termos que fazem parte de sua práxis. Dito de outro modo, codificamos a informação documentária a partir de referências que serão reconhecida, decodificadas e interpretadas pelo usuário no processo de semiose documentária.

A garantia de qualidade na representação e da recuperação da informação depende da efetiva a codificação e a decodificação da informação documentária, mas não está restrita a compreensão de sua estrutura, ela deve considerar também algumas tensões dialéticas entre forças contrárias presentes quando do gerenciamento e manutenção da linguagem documentária, como por exemplo, a **conservação x mudança** e **consenso x especificidade** (LIMA, 2004).

A **conservação** assegura o entendimento entre os sujeitos, e a **mutação**, satisfaz as necessidades de mudança da sociedade, ou seja, no processo de gerenciamento e manutenção de uma linguagem documentária deve-se manter pontos de acesso de sistemas anteriores ao mesmo tempo em que deve possibilitar a inclusão de novos pontos de acesso (LIMA, 2004).

A linguagem documentária deve atender também tanto ao **consenso** quanto a **especificidade**, isto é, de um lado atender a sua característica de ser institucional, pois sempre é construída para ser utilizada em função dos objetivos de determinada instituição assegurando a intercompreensão dos sujeitos e, de outro lado, oferecer elementos de uma experiência específica a cada usuário (LIMA, 2004).

A necessidade da linguagem documentária em garantir o consenso ao mesmo tempo em que atende a especificidade exige de seus descritores uma precisão e coerência que só é alcançada a partir da compilação de enunciados verdadeiros que vão explicitar seu significado e tornar explícitas as relações semânticas que articulam sua rede conceitual. Em outras palavras, é necessário identificar as características que formam o conceito, cuja designação (termo) é referente para o descritor que

atribuído a um documento constitui a informação documentária. Enfim, o sentido da informação documentária é expresso pela definição do conceito que ela denomina.

Por sua vez, a definição do conceito reúne atributos, ou características, que permitem estabelecer as categorias de uma linguagem documentária. Essas categorias se definem pelo traço comum presente a toda uma classe de conceitos/termos/descritores que por isso mesmo estão associados. O limite da cada categoria se estabelece pelos traços específicos que permitem individualizar cada conceito/termo/descritor estabelecendo a disjunção entre as informações documentárias elaboradas.

Aqui incluímos no percurso a noção de classe semântica que também foi abordada por Gardin (1966) quando na descrição do léxico documentário ao se referir à necessidade de organizar os seus termos de modo a explicitar as relações hierárquicas existentes entre eles, seja por afinidade ou diferença semântica.

Determinar o campo semântico em linguística é, segundo os pressupostos epistemológicos, procurar descobrir a estrutura de um domínio dado, ou propor a ele uma estruturação (Dubois, et al, 2011). Trier, citado por Lopes (1987), observa que as unidades léxicas de uma língua encontram-se organizadas em grupos estruturados de tal modo que, cada unidade fica ali definida pelo lugar que ocupa respectivamente à posição das demais. Assim, inferimos que o sentido de um conceito/termo/descritor é especificado por sua semelhança e sua diferença em relação aos outros elementos relevantes do campo semântico, pois uma palavra só adquire seu significado por oposição as outras unidades do mesmo campo (Germain, 1981).

Genouvrier e Peytard (1974) no que se refere ao campo semântico, o definem como o conjunto de empregos de uma palavra (ou sintagma, ou lexia) onde e pelos quais a palavra adquire uma carga semântica específica e a delimitação destes empregos se daria pelo levantamento de todos os contextos imediatos que a palavra recebe num texto dado.

Segundo Hernando Cuadrado (1995) a condição mínima para que as palavras pertençam a um mesmo campo é que tenham um traço significativo (sema) comum (quanto maior o número de semas tanto mais coerente será o campo semântico e no geral um menor número de palavras o integra); uma palavra pode tomar parte de todos os campos semânticos que se construam sobre qualquer dos traços significativos que se descubram nela; quando uma palavra possui vários significados, cada um deles pertence a um campo semântico distinto.

Por exemplo, ao mesmo tempo em que podemos identificar com pertencentes ao campo semântico da palavra *mesa,* devido à característica comum *"objeto que permite reunir em torno de"* as seguintes palavras: *mesa de refeições, mesa redonda*, *mesa da Assembléia*, *mesa de operação*, cada uma poderá fazer parte de uma categoria diferente, em uma linguagem documentária, devido à características delimitadoras como: "para refeição"; "para debate"; pra cirurgia", etc. Em outro caso, a palavra *lodo* por ser incluída em diferentes categorias de uma linguagem documentária a partir de características delimitadoras, a saber, *ser uma matéria prima; ser um produto, ser um reagente.*

No momento em que se discute a interoperabilidade entre as linguagens documentárias na web semântica acreditamos ser necessário aprofundar as discussões sobre o mapeamento dos campos semânticos que constituem um domínio, pois só a partir das características do conceito elencadas nas definições dos termos, que vão servir de referente e permitir a contextualização dos descritores das linguagens documentárias, será possível tornar mais eficaz a representação e a recuperação da informação.

Referências

CABRÉ, M. T. La terminologia hoy: concepciones, tendências y aplicaciones. **Ciência da Informação**, Brasília, v.24, n.3, p.289-298, set./dez. 1995.

CROS, R.C.; GARDIN, J-C; LEVY, F. **L' automatisation des recherché documentaries: un modele general: le Syntol**. Paris: Gauthiers-Villars, 1968.

DUBOIS, J. ET AL. ***Dicionário de linguística***. São Paulo: Cultrix, 2011.

GARDIN, J. C. Éléments d'un modele pour La description des lexiques documentaires. ***Bulletindes Bibliothèques de France (en ligne)****. n.*5, 1966. Disponível em http://bbf.enssib.fr/consulter/bbf-1966-05-071-001, Acessado em 20 jun. 2015.

GARDIN, J. C. ***La logique du plausible***. Paris: Editions de la Maison des Sciences de l´homme, 1981.

GENOUVRIER, E., & PEYTARD, J. ***Linguística e ensino do português***. Coimbra: Almedina, 1974.

GERMAIN, C. (1981). ***La sémantique fonctionnelle***. Paris: PUF, 1981.
Hernando Cuadrado, L. A. ***Introduccíon a la teoria y estrutura del lenguage***. Madrid : Editorial Verbum, 1995.

LARA, M. L. G. Conceitos de organização e representação do conhecimento na ótica das reflexões do Grupo Tema. **Inf. Inf.,** Londrina, v. 16. n. 3. p. 92 – 121, jan./ jun. 2011.

LARA, M. L. G. É possível falar em signo e semiose documentária? **Enc. Bibli: R. Eletr. Bibliotecon. Ci. Inf**., Florianópolis, 2° número esp., 2º sem 2006

LARA, M. L. G. ***Representação e linguagens documentárias: bases teórico-metodológicas.*** (doutorado), Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 1999.

LIMA, V.M.A. ***Da classificação do conhecimento científico aos sistemas de recuperação de informação: enunciação de codificação e enunciação de decodificação da informação documentária.*** (doutorado), Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo

LIMA, V.M.A. ***Terminologia, comunicação e representação documentária***. 1998. Dissertação (Mestrado) Escola de Comunicações e Artes, Universidade de São Paulo, 1998

LOPES, E. ***Fundamentos da linguística contemporânea***. São Paulo: Cultrix, 1987. SMIT, J. W. Informação. In: LIMA, Y.D.; SMIT, J.W. (Org.) **Organização de arquivos**. São Paulo: IEB/ECA, 2000. p.19-31

PEIRCE, C.S. ***Semiótica.*** São Paulo: Perspectiva, 1977.

TÁLAMO, M.F.G.M.; LARA, M.L.G.; KOBASHI, N.Y. & LIMA, V.M.A. Instrumentos de controle terminológico: limites e funções. ***Actas do II Simpósio Latino-americano de Terminologia****,* 1990. Disponível em: <http://www.riterm.net/actes/2simposio/indice90.htm>. Acesso em 20 de junho de 2015.

VOGEL, M.J.M. A influência de Jean-Claude Gardin e a linha francesa na evolução do conceito de linguagem documentária. ***Perspectivas em Ciência da Informação***, v.14 , número especial, p.80-92, 2009.

Ordenação de documentos: fundamentos e relações com a classificação bibliográfica

Cristina Dotta Ortega
Universidade Federal de Minas Gerais
ortega@eci.ufmg.br

1 Introdução

A ordenação de documentos é atividade que permite a proposição de um arranjo para a coleção, pertinente aos objetivos institucionais, às características dos documentos e às atividades do público-alvo. Esse arranjo tem por funções fornecer um modo de leitura da coleção que se realiza pela circulação no espaço, possibilitar o acesso aos documentos por meio de sua localização a partir de códigos indicados em registros de bases de dados e nos documentos dispostos na coleção e viabilizar a gestão da coleção no que tange a políticas de seleção, otimização do espaço e preservação dos documentos. Essas funções podem ser operacionalizadas por um plano de ordenação, instrumento que fornece um mapa estruturado para a disposição espacial da coleção, incluindo os códigos a serem utilizados para os documentos.

Atividade realizada em abordagem arquivística, museológica e bibliográfica, a ordenação de documentos será aqui tratada apenas quanto à última. Embora difícil de ser definida, segundo nos parece, podemos dizer que a abordagem bibliográfica da informação contempla: informação para construção de conhecimento, necessária a atividades educacionais, científicas e profissionais em geral; informação para fruição ou experiência estética; e informação para fins utilitários, relativa ao acesso a serviços ou atividades de entretenimento, educação, cultura, saúde e direitos civis em geral. A perspectiva bibliográfica envolve também a experiência sensorial a partir de objetos

em exposições ou atividades similares, que não trataremos aqui, priorizando o caráter cognitivo das mensagens produzidas sobre objetos.

Os problemas que justificam o tema deste trabalho são descritos a seguir.

A despeito da necessidade de adequação de procedimentos a seus contextos visando a comunicação com públicos, um modo de realizar a atividade de ordenação de documentos predomina frente a outros. Trata-se do número de chamada, criado com a função de código de localização do documento em uma coleção. O número de chamada pertence ao conjunto de soluções de processamento técnico de coleções de biblioteca que compõe o que chamamos de modelo anglo-americano. A adoção extensiva do número de chamada pela comunidade de bibliotecas constitui-se pelo uso da Classificação Decimal de Dewey (CDD) ou da Classificação Decimal Universal (CDU) para a notação classificatória, cuja função é agrupar documentos, seguida do número correspondente ao nome do autor, em geral o ponto de acesso principal da ficha do documento, segundo as Regras Anglo-Americanas de Catalogação (AACR2), com uso da tabela Cutter de nomes de autor, com a função de individualizar o documento. O modelo do número de chamada ganhou maior legitimidade e capacidade de disseminação no Brasil por compor os conteúdos dos cursos de graduação em Biblioteconomia. No entanto, em geral, nos programas de ensino não são aprofundados os aspectos históricos, teóricos e metodológicos que fundamentam o número de chamada. Mais que isso, um recorte ideal partiria do conceito de código de localização do documento, deste modo, passível de adequação a diversos contextos.

O modelo único faz prevalecer o uso dos tradicionais sistemas de classificação bibliográfica citados, a CDD e a CDU. Enquanto instrumentos componentes de um modelo, vários pontos merecem aprofundamento. Primeiramente, está a questão de que sistemas de classificação bibliográfica não permitem a ordenação pois não realizam sozinhos a individualização do documento em uma coleção. Além disso, a ordenação não é realizada apenas pelo método classificatório, podendo ocorrer pelo método sequencial, ou por ambos. Assim, uma operação é tomada pela outra, ou seja, a classificação faz as vezes da ordenação. O segundo ponto é o de que os sistemas de classificação bibliográfica são estudados na perspectiva da representação temática dos documentos, juntamente com as listas de cabeçalhos de assunto e os tesauros. No entanto, os atributos dos documentos pertinentes a uma disposição física dos mesmos, os quais estão claramente indicados nos tradicionais sistemas de

classificação bibliográfica e no modo como são utilizados, não se referem apenas a assuntos. Uma terceira questão é a compreensão de que os tesauros seriam mais relevantes por atenderem à recuperação da informação, frente a um peso menor que os sistemas de classificação bibliográfica exerceriam como ordenadores de documentos. Como uma atividade não se sobrepõe à especificidade da outra, a ordenação continuou necessária mas, tanto o processo em si, quanto os instrumentos adotados para ele, não foram suficientemente estudados. Como quarta e última questão, apontamos o problema da assimilação do conceito de organização da informação pelo de classificação. Embora, em sentido amplo, todos os procedimentos de organização da informação impliquem operações cognitivas classificatórias, eles não se restringem a estas operações.

Pelo dito anteriormente, constatamos a existência de lacunas em Organização da Informação sobre processos e instrumentos de ordenação de caráter classificatório, mas não somente eles, que dificultam o desenvolvimento da mesma já que, a compreensão de um processo ocorre na medida em que o mesmo é devidamente explicitado, assim como, sua relação com os demais processos.

Desse modo, o objetivo deste trabalho é o de explorar fundamentos à ordenação bibliográfica de documentos, por meio do estudo da especificidade do processo e de seus instrumentos e quanto às suas relações com o processo de classificação bibliográfica.

A metodologia incluiu leitura e seleção de textos sobre conceitos fundamentadores do processo de ordenação de documentos e, de sua operacionalização, integrados a demais processos. Foram selecionados e adotados para a pesquisa textos em português (do Brasil e de Portugal), espanhol, inglês e francês. Adotamos como referência principal a produção francesa contemporânea sobre o tema, desenvolvida em torno do termo ordenação (*classement*, na forma original em francês), por sua maior generalização sobre os processos, instrumentos e produtos que compõem a atividade, mas também por ser proposta em uso nas instituições francesas. Quanto aos métodos de ordenação – sequencial e classificatório –, uma breve sistematização é proposta a partir de indicações dispersas da literatura, basicamente, Salvan (1962), Jouguelet (1997), Maniez (2002), Carrión Gútiez (2002), Calenge (2010) e Combot (2011). O estudo dos instrumentos classificatórios da ordenação e dos sistemas tradicionais de classificação bibliográfica pautou-se especialmente em Gardin e Coyuad que, a partir de textos próprios, ambos

publicados em 1966, na França, trataram da abordagem linguística como fundamentação a esquemas classificatórios, a partir da ideia da unidimensionalidade que é intrínseca aos mesmos. Baseada em grande medida nesta abordagem, no contexto dos estudos do Grupo Temma, da Escola de Comunicações e Artes (ECA) da USP, Lara (1999, 2001 e 2002) discorre sobre os aspectos linguísticos dos sistemas tradicionais de classificação bibliográfica, além de observar especificamente seus aspectos pragmáticos.

Como estrutura do trabalho, no tópico 1 contextualizamos a ordenação de documentos como processo de organização da informação. No tópico 2 desenvolvemos uma conceituação do processo de ordenação de documentos, por meio de seus processos relacionados, seus métodos, instrumentos e produtos. No tópico 3, discutimos a especificidade da classificação bibliográfica, de modo a avançar na compreensão da ordenação de documentos, observando a singularidade de cada uma e as interconexões entre elas. No tópico 4, discorremos sobre os fundamentos linguísticos dos instrumentos classificatórios de ordenação, analisando, com base neles, os sistemas de classificação bibliográfica mais conhecidos. Apresentamos ao final algumas considerações sobre o tema.

2 Contextualizando a ordenação de documentos

Como atividade documentária, a ordenação de documentos é realizada no contexto das ações de mediação da informação. A apropriação, que se objetiva por meio dessas ações de mediação, devem ser abordadas, segundo Perrotti e Pieruccini (2007, p. 60, 88), num quadro amplo de questionamentos que envolve a informação e seus dispositivos (bibliotecas, centros e núcleos de documentação, de informação, de memória, dentre outros), em suas múltiplas interações e ângulos. Falamos em mediação entre objetos e pessoas no sentido do enfrentamento do objeto, segundo Meneses (1994, p. 30), que conduz a sua transformação em documento, e da observação de pessoas como usuárias potenciais de informação. A mediação é processo simbólico que opera sobre objetos e pessoas na perspectiva de documentos e usuários, de modo que, tanto documentos quanto usuários definem-se situacionalmente.

As pessoas a que chamamos usuários são aquelas abordadas em alguma dimensão de suas vidas a partir da qual certas informações podem ser utilizadas com

proveito, seja científico, utilitário, ou outro. Embora todas as pessoas sejam usuárias de informação, já que a necessidade e o uso de informação são inerentes ao ser humano, o campo trabalha com estas necessidades de informação de modo particular, buscando satisfazê-las mas, também, sensibilizando públicos. Assim, trata-se antes de público, que de usuários, no sentido de pessoas que são destinatárias da produção de mensagens sobre objetos.

Documentos são objetos abordados informacionalmente. O documento se define por uma instância física (seu suporte material) e uma instância simbólica (informativa, conteudística), sendo que esta última é sempre uma atribuição, ou seja, não é inerente ao objeto. Assim, a dimensão informativa do documento não é prévia, não existe *a priori* ou *in natura.* É a instância definidora da condição de ser do documento, é construída no momento da interpretação ou do 'fazer-se um documento', ainda que a dimensão física seja necessária para tal. Para Meyriat (1981, p. 54) e Sagredo Fernández e Izquierdo Arroyo (1983, p. 265), 'ser documento' depende de que o mesmo seja usado como tal. Meyriat (1981, p. 51) afirma também que o documento pode ser definido como um objeto de suporte da informação, que serve para comunicar, e que é durável, de modo que a comunicação possa se repetir.

Se o documento é objeto construído, o documento bibliográfico, que tratamos neste texto, também não é objeto pré-existente, mas depende de ações humanas de significação que o tornem como tal. É preciso superar a crença de que o documento bibliográfico é previamente definido por sua tipologia, caso do livro e do periódico, embora estes tipos sejam recorrentes em sistemas de informação bibliográfica como bibliotecas, centros de documentação e bases de dados especializadas, entre outros, por apresentarem conteúdos científicos ou estéticos. Outra revisão que se faz urgente é a da ideia de que o suporte alteraria a tipologia documental; ao contrário, um livro, ao apresentar um certo tipo de mensagem condicionada por uma forma de organização de conteúdos que lhe caracteriza, não deixa de ser um livro caso tenha sido produzido em papel ou em meio eletrônico.

Considerando que nenhum objeto é por si mesmo um documento, o processo de transformação do primeiro no segundo é decorrente das atividades documentárias que realizam a produção de mensagens sobre objetos frente a usuários. Desse modo, sob o ponto de vista do campo que nos ocupa, um objeto faz-se documento a partir dessas atividades.

Outro momento, no entanto, é aquele da interpretação realizada pelos usuários: essa interpretação é influenciada pela produção de mensagens sobre objetos, mas não é determinada por ela. Não há relação de correspondência direta entre as mensagens sobre objetos e a interpretação das mesmas mas, a observação das diversas apropriações realizadas pelos usuários (ou não realizadas) contribui como modo de avaliar e rever as atividades documentárias.

Por meio das atividades documentárias são produzidos sistemas de informação documentária ou sistemas documentários. Dentre eles, tratamos dos sistemas de informação bibliográfica, como dissemos. Inicialmente caracterizados pelas bibliografias e bibliotecas com seus acervos e catálogos, hoje apresentam uma amplitude maior de variações a depender das tipologias e suportes documentais, dos conteúdos dos documentos e dos públicos e atividades em que estes se inscrevem. Temos, assim, repertórios bibliográficos diversos como bibliografias, catálogos de bibliotecas, bases de dados, portais de informação e estruturas hierárquicas para navegação e acesso de recursos *online*, entre outros.

Em termos gerais, podemos considerar como atividades documentárias: identificação das necessidades de informação; seleção; coleta; representação; ordenação; preservação; e elaboração de produtos e serviços de informação, exposições e ações educativas e culturais.

Para a identificação das necessidades de informação, é realizado diagnóstico do fluxo de produção e uso de informação no contexto das atividades que estão em questão. A partir do diagnóstico, é possível obter critérios que subsidiem a seleção de documentos.

O produto da seleção de documentos é a coleção de documentos, a qual é determinante na constituição do sistema por fornecer-lhe personalidade. Os documentos da ‘coleção’ são, ou não, coletados. Os documentos podem não ser coletados quando há a indicação do seu endereço eletrônico ou do nome e localização da pessoa ou instituição depositária dos mesmos; há, ainda, bases de dados que não apresentam a informação de localização para acesso, como no caso das bases de dados científicas, em especial no período anterior às publicações de revistas em meio eletrônico. Quando os documentos são coletados, é realizada sua ordenação por meio de códigos, ou não; quando realizada, a ordenação dos documentos é a última etapa da organização da informação, pois é aquela em que os documentos tornam-se finalizados para acesso e uso pelo usuário.

A representação dos documentos ocorre nos próprios documentos – no caso dos códigos a eles atribuídos para sua ordenação em um certo espaço – e/ou em repertórios bibliográficos, como bases de dados, na forma de registros –, caso em que ocorre a ordenação dos seus pontos de acesso.

As atividades de preservação dos documentos são necessárias para fornecer possibilidades de usos posteriores.

Em geral, em função das atividades anteriores, são realizados os serviços de informação e demais ações e produtos relacionados. como: a busca sistêmica e contextualizada da informação em sistemas diversos ou em nenhum deles; o acesso ao documento (a exceção das bases de dados cadastrais, nas quais o registro deve responder ele mesmo à questão da busca); a disseminação seletiva da informação; o apoio para o uso da informação e a formação de usuários; a produção de exposições; e as ações educativas e culturais diversas para sensibilização de públicos.

Dentre as atividades documentárias citadas, aquelas que constituem especificamente a organização da informação são indicadas a seguir:

- seleção de documentos;
- coleta de documentos;
- representação de documentos na forma de registros componentes de repertórios bibliográficos, por meio dos processos de descrição, indexação e elaboração de resumos, e da escolha e forma de pontos de acesso, e na forma de códigos de localização de documentos, muitas vezes com uso do processo de classificação bibliográfica; e
- ordenação de documentos e ordenação de metadados de documentos, a partir de códigos dos documentos, no primeiro caso, e de pontos de acesso a registros de repertórios bibliográficos diversos, no segundo.

Como podemos ver, a ordenação é processo mais abrangente, envolvendo mais que a atividade de alocação física de documentos. A ordenação em sentido amplo foi definida, já em 1962, mas muito provavelmente antes disso, como atividade que envolve documentos, fichas de um catálogo ou referências de uma bibliografia ou de um catálogo impresso (SALVAN, 1962, p. 1). Assim, no conjunto dos procedimentos que compõem a organização da informação, a ordenação deve ser compreendida a partir de suas duas facetas: a ordenação de documentos e a ordenação dos metadados dos documentos. No primeiro caso, tem-se a disposição física de documentos, como tratamos. O segundo caso faz menção aos metadados,

termo utilizado atualmente para indicar o conjunto das informações relativas a um documento na forma de um registro que compõe os diversos tipos de repertórios bibliográficos. Considerando a atividade no tempo, a ordenação de metadados refere-se à ordem atribuída às referências de bibliografias e seus índices de acesso, às fichas catalográficas segundo seus cabeçalhos, aos pontos de acesso dos índices das bases de dados de qualquer tipo e aos termos que compõem as hierarquias sobre recursos *online*. Assim, tanto a ordenação de documentos quanto a ordenação dos metadados dos documentos demonstram, simultaneamente, a anterioridade e a atualidade da atividade.

A atividade de ordenação apresenta funções de caráter comunicativo e gerencial, o segundo sempre voltado a dar conta do primeiro. Elegemos três funções como principais: modo de leitura, localização dos documentos e gestão da coleção, como segue.

A ordenação envolve produção de mensagens, neste caso, modos de leitura da coleção de documentos enquanto propostas orientadoras da sua exploração pela circulação no espaço, seja o tradicional ou o eletrônico. Estas propostas de modos de leitura implicam a apresentação dos documentos em uma determinada perspectiva, ou seja, segundo certos aspectos e não outros. Como modo de orientação para certos percursos e olhares sobre a coleção, é possível saber o que há nela, buscar pelos documentos de interesse e identificar outros cuja relação foi estabelecida na ordenação. As exposições e atividades similares também estão incluídas nesta função a seu modo.

Outra função da ordenação de documentos é a de localização dos mesmos em uma determinada coleção. Um código ou um endereço eletrônico atribuído a cada documento singulariza o mesmo frente aos demais em uma coleção, possibilitando o acesso, por realizar a relação entre o registro do documento na base de dados e sua localização em um determinado espaço tradicional ou eletrônico. Em uma coleção tradicionalmente disposta, o código permite que o documento seja retirado e guardado novamente segundo a ordenação proposta. No caso de endereços eletrônicos de documentos para localização, uma proposta de arranjo propriamente dita não está em questão, a não ser quando há estruturação hierárquica para navegação e acesso. Desse modo, os pontos de acesso dos diversos repertórios bibliográficos remetem aos registros a que se referem que, por sua vez, indicam os documentos correspondentes, quando é o caso, como vimos.

A atividade é adotada também para a gestão das coleções no que se refere a uma política de seleção para aquisição e descarte de documentos e ao planejamento dos espaços para sua alocação. No que tange ao melhor uso do espaço, estão em questão a previsão de crescimento da coleção e as condições e os modos mais adequados de armazenamento segundo suporte e tipologia documentais, tanto para fins de preservação quanto para fins de acesso.

A despeito da usual ênfase gerencial atribuída às atividades de seleção e de estudos de usuários, todas as atividades documentárias devem ser devidamente geridas, portanto, abordadas de modo específico, mas também articulado, em um processo de gestão integrada, segundo métodos e ferramentas que permitem uso racional, econômico e eficaz.

Como toda atividade de organização da informação, a ordenação deve ser orientada pelos objetivos institucionais, pelas características da coleção de documentos e pelas atividades do público-alvo para as quais certas informações podem ser utilizadas.

3 Conceitos e métodos de ordenação de documentos

Para explorar o tema, selecionamos terminologia sobre ordenação de documentos, que sistematizamos a seguir, oriunda da literatura produzida na França, em função do grau de generalização que apresenta. É digno de nota, também, o fato de tratar-se de modelo aplicado. Como cada idioma representa sua própria visão de mundo, e estas diferenças idiomáticas são pouco claras quanto menor for a consolidação teórica do campo em questão, os termos franceses usados como referência não apresentam correspondência direta com os termos adotados no Brasil, a despeito de refletirem algumas soluções significativas da prática profissional brasileira.

Identificamos no livro "*Mettre en oeuvre un plan de classement*", coordenado por Calenge (2010), o material mais atual e consolidado sobre o tema da ordenação de documentos, incluindo conceitos, questões e diversos relatos de experiências. Os seguintes termos foram retirados do glossário deste livro (CALENGE, 2010, p. 191-196): ordenação, quadro de ordenação, plano de ordenação, códigos validados, manual de codificação, código, codificação, etiquetagem, recodificação e tabela de correspondência. Foram usados também verbetes de dicionários especializados,

como segue: Jouguelet (1997, p. 135-136: ordenação), Boulogne (2005, p. 67: código - *code*; p. 59: código - *cote*; p. 56: ordenação) e Cacaly et al. (2008, p. 55: código - *code*), além do material didático de Combot (2011, p. 7: plano de ordenação; p. 29: cotação com símbolos, com cores).

A **ordenação** (*classement*) **de documentos** é definida como o processo de arranjo de uma coleção ou de parte de uma coleção de documentos em móveis como prateleiras ou arquivos, decorrente de operações baseadas em uma ordem pré-estabelecida a partir de um determinado critério de interesse.

Sob o ponto de vista instrumental, para realizar a atividade pode-se dispor de um **quadro de ordenação** (*cadre de classement*), o qual apresenta o sistema geral da ordenação de documentos na forma de uma arquitetura do modo de ordenação que é proposto. O quadro de ordenação seria um mapa ou um esquema do modo de ordenação dos documentos, desenvolvido especialmente para implementação em uma coleção determinada.

Por sua vez, o instrumento que efetivamente permite operacionalizar a atividade segundo sua própria sistemática é o **plano de ordenação** (*plan de classement*), que citamos inicialmente. Trata-se de um texto que descreve a organização, as regras de ordenação e os modos de acesso ao conjunto de documentos que constitui uma coleção. O plano predetermina a organização do espaço documental que é especificada por meio de outro instrumento, o manual de codificação, composto pelos códigos validados (descritos a seguir). O plano de ordenação deve ser passível de ampliação e assegurar a preservação dos documentos. Deve atribuir a cada documento um lugar único e permitir encontrar fácil e rapidamente um documento, assim como, guardá-lo novamente no lugar, possibilitando a manutenção da ordem inicial com baixo custo.

Os **códigos validados** (*cotes validées*) são aqueles determinados por um plano de ordenação e listados em um manual de codificação. Estes códigos respondem a uma organização espacial e intelectual do espaço documental que é determinada *a priori*, como dissemos. O **manual de codificação** (*manuel de cotation*) é a lista do conjunto de códigos validados em uma coleção, os quais são seguidos de notas de uso, remissivas, detalhamentos do uso, etc., podendo conter também um índice. Desse modo, o termo códigos validados indica que se trata de códigos construídos previamente e segundo uma estrutura ou mapa próprio de organização, à diferença do uso adaptado de códigos de sistemas de classificação bibliográfica, ou

ainda, do uso de códigos destes sistemas, documento a documento, na ausência de adaptação, como tratamos à frente.

Um código (*code*) é uma cadeia de caracteres breve, simples e estruturada, cuja construção é baseada sobre um conjunto de regras, tendo por objetivo a codificação. A capacidade teórica do código deve corresponder às necessidades da codificação, atendendo a características de univocidade e estabilidade. Quanto à ordenação de documentos, o **código** (*cote*) é um dado de localização de um documento em uma coleção. O código do documento permite identificá-lo e ordená-lo no conjunto de documentos da coleção. Os seus elementos constituintes apresentam sucessão estruturada que fornece a orientação a ser dada à ordenação dos documentos e à sua localização. O código é geralmente representado por um conjunto de letras ou números, que será reproduzido no documento e em todas as suas 'imagens' (registro de entrada, catálogos etc.), o que permite que o documento seja encontrado onde foi alocado. Por estes motivos, os códigos devem ser mnemônicos.

A **codificação** (*cotation*) é a operação de elaboração de um código para um documento pertencente a uma coleção, que determina a posição ou localização do documento na mesma. Envolve uma etapa propriamente intelectual relativa à elaboração do código, seguida de uma etapa marcadamente material que se dá com a inscrição do código no registro do documento na base de dados e a produção de uma etiqueta e a colagem desta no documento. A **etiquetagem** (*rondage*), por sua vez, é a operação que consiste em fixar sobre o documento a etiqueta com o código. A etiqueta com o código pode ser produzida manualmente ou por computador. Neste caso, lembramos que a etiqueta pode ser produzida diretamente do código que foi inserido no registro da base de dados.

A **recodificação** (*recotation*) ocorre quando se opta por um novo modo de ordenação que conduz à mudança de lugar dos documentos e à atribuição de outro código conforme o novo quadro de ordenação escolhido. Do mesmo modo que o processo de codificação, há a etapa intelectual de atribuição do código e a etapa material de colocar o código na parte exterior do documento. No processo de recodificação, é funcional estabelecer uma **tabela de correspondência** sobre a qual são colocados em paralelo os novos códigos adotados face aos antigos códigos existentes.

Em qualquer caso em que haja documentos ordenados no espaço de uma biblioteca ou similar, podem ser produzidas legendas, as quais são fixadas às

prateleiras, estantes e demais móveis em que estejam dispostos os documentos, de modo a orientar a circulação pelo espaço. Também para contribuir para a localização espacial dos conjuntos documentais pelo usuário e para uma circulação proveitosa por ele, vários outros recursos são utilizados, como o das cores acopladas aos códigos e, de modo correspondente, às legendas. É necessário observar que as cores não exercem a função de classificação, como muitas vezes se afirma no Brasil, mas função de sinalização. De fato, as cores, elas mesmas, não têm potencial para a indicação de atributos ou características dos documentos, a exemplos dos termos e dos números relacionados aos mesmo em uma estrutura hierárquica. Assim, é importante reforçar a incongruência do termo 'classificação por cores', a exemplo de Combot (2011, p. 29) que trata, de modo mais apropriado, de **codificação com símbolos ou cores** (*cotation avec des symboles, des couleurs*). Segundo ele, faz-se a ordenação dos documentos segundo um critério qualquer e se atribui a esse critério uma marca, que não é alfabética nem numérica, como uma cor ou um pictograma, que é colada no item por meio de uma etiqueta ou similar. Combot considera o método adequado para algumas bibliotecas, como aquelas para crianças em que uma cor por nível de leitura ou por classe, por exemplo, é bastante útil. A constatação da pertinência do termo 'codificação com símbolos' e da impropriedade do termo 'classificação por cores' deve contribuir para a intensificação de estudos sobre a sinalização em bibliotecas e espaços similares.

É no caso de livrarias, de alguns acervos pessoais, de obras raras e de documentos frágeis que códigos não são adotados, sob pena de que sua materialidade original seja alterada. Eventualmente, para estes casos, tiras com os códigos são inseridas nos livros ou outros documentos, sem que sejam neles aderidas.

Assim, na atividade de ordenação em sentido amplo, as formas de representação e acesso podem ser códigos em etiquetas coladas aos documentos, tiras dispostas nos documentos, endereços eletrônicos dos documentos e metadados em geral como cabeçalhos de fichas catalográficas, pontos de acesso às bases de dados, categorias das estruturas de recursos *online*.

Em organização da informação, consideramos pertinente sistematização baseada em processos, instrumentos e produtos, a qual adotamos para tratar dos conceitos básicos da ordenação de documentos. Deste modo, a terminologia apresentada está estruturada a partir do processo de ordenação de documentos, o

qual envolve os **processos** de classificação bibliográfica, codificação, etiquetagem e recodificação. Os **instrumentos** adotados para a realização da ordenação são o quadro de ordenação, o plano de ordenação, os códigos validados, o manual de codificação e a tabela de correspondência. Os **produtos** resultantes são os documentos ordenados.

Quanto aos **métodos** adotados na ordenação, podemos falar em sequencial e em classificatório ou sistemático. A partir de indicações esparsas da literatura (SALVAN, 1962, JOUGUELET, 1997, CARRIÓN GÚTIEZ, 2002, MANIEZ, 2002, CALENGE, 2010 e COMBOT, 2011), propomos breve sistematização dos métodos de ordenação de documentos, como segue.

Questão primordial é a de que o tipo de atributo do documento considerado é o critério norteador do método. Os atributos de um documento, sempre situacionais, é que o definem como tal, como tratamos anteriormente. O método sequencial pode ser realizado de dois modos. No método sequencial contínuo, cada documento é disposto um seguido do outro, baseado em alguma cronologia de eventos, a qual exige codificação numérica; neste caso, os atributos em questão são a ordem de chegada, os fascículos de uma revista, a data de publicação. No método sequencial intercalado, cada novo documento pode ser inserido no meio daqueles já dispostos no espaço, a partir de algum de seus termos de identificação, como autor, título, série, para o qual a codificação alfabética é pertinente. O método classificatório ou sistemático, por sua vez, implica organização hierárquica de suas unidades, em uma estrutura unidimensional, por meio de agrupamentos baseados em classes e sub-classes. O primeiro método – sequencial contínuo – apresenta localização fixa, diferente dos outros dois, cuja intercalação conduz a uma localização relativa dos documentos.

Algumas situações a considerar, no que tange à espacialidade para uma ordenação de documentos, são as que seguem.

A ordenação segundo o princípio de proveniência dos documentos ocorre nos casos em que coleções são doadas a bibliotecas, implicando uma ordenação em separado do restante da coleção da biblioteca e que considere o arranjo original. Trata-se do arranjo original dos documentos segundo seu funcionamento na organização ou na vida da pessoa a que pertencia, em referência a uma abordagem claramente arquivística.

Tem-se, ainda, a ordenação por atributos materiais (tamanho, suporte, cor da capa) ou tipológicos do documento, muitas vezes também levando a uma separação dos documentos de um certo tipo ou suporte do restante da coleção em função da exigência de acesso rápido (caso das obras de referência alocadas na entrada de uma biblioteca) ou de preservação (condições especiais de temperatura e umidade, como no caso dos audiovisuais, ou necessidade de localização fixa para documentos frágeis).

Outra situação relativa a espaços é o das redes de informação, como as de informação científica, em que a alocação dos documentos ocorre em cada biblioteca que compõe a rede, de tal modo que o código do documento no registro das bases de dados deve ser antecedido pelo código da biblioteca que é proprietária do documento.

Calenge (2010, p. 191-196) fala que os códigos podem estar baseados: em aspectos como o formato ou a ordem de entrada, usuais em acervos reservados; sobre um sistema de classificação adaptado à coleção, caso em que a lista de códigos validados é construída a partir de um sistema de classificação pré-existente; ou sobre um plano de ordenação, o qual inclui códigos validados construídos localmente. Constatamos que, em todos os casos, a maior parte dos autores e os exemplos da prática profissional indicam que um plano de ordenação está em questão. No entanto, para Calenge, apenas no último caso haveria um plano de ordenação propriamente dito.

Calenge (2010, p. 193) trata ainda de outra situação em que os códigos são produzidos a partir de um sistema de classificação pré-existente: neste caso a escolha do código se efetua no momento da análise de cada documento e segundo o sistema adotado. Para ele, esses códigos não são chamados de códigos validados. De fato, por este procedimento não ocorre a adaptação de um sistema de classificação para a produção prévia e local dos códigos. Assim, a estrutura classificatória do sistema independe daquela que seria mais adequada à coleção de documentos. A situação citada é usual nas bibliotecas brasileiras, em especial, aquelas que adotam o número de chamada com uso da CDD ou da CDU e da tabela Cutter de nomes de autor, segundo o modelo anglo-americano, de que falamos. Por sua vez, na França, segundo Combot (2011, p. 30), para a produção dos códigos, utiliza-se em geral uma notação da CDD, seguida das três primeiras letras do nome do autor. Ele explica que não se ensina, no país, a tabela de nomes de autor de Cutter e pergunta como fazer para diferenciar autores com as mesmas letras.

Quanto ao modelo francês apresentado, Calenge (2010, p. 13-15) informa que os chamados códigos validados passaram a ser utilizados na *Bibliothèque Public d'Information*, do *Centre Pompidou*, em Paris, na forma de uma lista restritiva de notações da CDU, em função do uso descontrolado de códigos diversos que ocorria, de tal modo que o princípio do plano de ordenação nasceu a partir da CDU. A motivação, segundo Calenge, foi a busca por legibilidade e racionalidade dos espaços. Para ele, o plano de ordenação é resultado do trabalho de agrupamentos de documentos por meio da produção de códigos de modo estruturado, segundo as massas documentais em jogo, os públicos em causa e os limites arquiteturais. O documento final representa a formalização do sistema de endereçamento físico dos documentos proposto, em especial, para acervos de livre acesso. Para Calenge, o plano de ordenação é uma ferramenta que torna visível a política documentária produzida para um público específico. Ao mesmo tempo, configura-se como uma ferramenta de gestão que, como tal, caracteriza-se por procedimentos, portanto, uma ferramenta de trabalho cotidiano do pessoal envolvido que permite a cada um "adaptar-se, discutir e contestar" (p. 15). Disto observamos que, assim como no caso da compreensão, relativamente consensual, da necessidade de construção de tesauros próprios a cada sistema, o plano de ordenação que, por sua própria origem e definição, é local, mostra-se desejável.

A distinção dos modelos aponta para soluções com características próprias. De um lado, o modelo francês, adota o plano de ordenação como instrumento que mapeia e operacionaliza a ordenação dos documentos, servindo como instrumento para a leitura da coleção e circulação pela mesma, acesso aos itens, seleção e descarte, preservação e otimização do espaço. De outro lado, o modelo anglo-americano tem no catálogo topográfico um instrumento de uso interno para o controle dos itens da coleção, em especial aquela que é disponível diretamente ao público. Com a informatização dos processos de gestão das coleções de bibliotecas, módulos de controle dos itens da coleção foram desenvolvidos, possibilitando recursos mais potentes de gestão como aqueles que contabilizam itens existentes, aquisições, descartes etc. Este módulo não funciona, no entanto, como instrumento para a ordenação dos documentos, segundo as funções que identificamos para a atividade. Por outro lado, as soluções apresentadas para a constituição do número de chamada são bastante completas para a individualização, em especial, de livros.

Seria válido retroceder no tempo em busca da compreensão sobre a construção dessas propostas, como realiza Weitzel (2012). A partir da análise de obras de autores do século XIX que tratam de desenvolvimento de coleções, a autora constatou que os mesmos pautavam-se fortemente pelas ideias de política, planejamento, plano e critérios para desenvolver coleções e, ainda, de estrutura para sua formação. Weitzel (p. 182) cita ainda os principais modelos atuais de estrutura para coleções conhecidos no Brasil, dentre eles, o Conspectus, da *International Federation of Library Associations and Institutions* (IFLA), em que os temas e os assuntos das coleções são determinados pela concentração de exemplares em classes e subclasses de um esquema de classificação. Assim, para uma consolidação do tema da ordenação de documentos, carecem de exploração, tanto aspectos históricos, quanto propostas contemporâneas.

4 Ordenação de documentos e classificação bibliográfica

A compreensão da atividade de ordenação de documentos demanda retomar o papel da classificação bibliográfica. Como vimos, a atividade de ordenação ou arranjo de documentos é usual no idioma francês na forma *classement*. Localizamos o uso do termo por Suzanne Briet, para quem (1951, p. 24) a ordenação de documentos não se faz da mesma maneira para a venda em uma livraria, para a exposição em um museu de arte e para a consulta em uma biblioteca especializada. Assim, Briet demonstra a amplitude do termo ordenação ao tratar da atividade segundo contextos variados. Atualmente, segundo Jouguelet (1997), o termo francês *classement* é adotado para indicar a **ordenação dos documentos** no espaço, ou seja, a operação material de colocar em ordem, alocando fisicamente os documentos, uns em relação aos outros, enquanto o termo *classification* refere-se à **classificação** no sentido de operação intelectual para a realização daquela. Essa distinção é recorrente na literatura contemporânea publicada na França.

Em uma abordagem arquivística, na França, segundo Hudon (2009, p. 70, 76-77), utiliza-se mais o termo plano de ordenação (*plan de classement*) que plano de classificação (*plan de classification*). Este instrumento apresenta uma estrutura essencialmente hierárquica, constituída pela enumeração de grandes funções e atividades administrativas em uma instituição, ou de atividades principais de um indivíduo, para a organização de documentos correspondentes às mesmas. O objetivo

é o de preservar o contexto de criação de um documento e, por consequência, determinar seu valor primário (valor administrativo, financeiro ou legal) e seu valor secundário (valor de testemunho histórico). Como podemos observar, a preferência pelo termo plano de ordenação parece emblemática da ideia de ordenação enquanto arranjo de documentos segundo algum princípio de interesse: o termo indica que o plano de ordenação em si não é de classificação, sendo este, de fato, o método adotado para aquele. A exemplo do livro citado, produzido em francês, no Canadá, usa-se o termo plano de classificação, forma que também é adotada no Brasil.

Em espanhol, no livro clássico "*Manual de Bibliotecas*", Carrión Gútiez (2002, p. 391-402) trata da atividade de ordenação de coleções (*ordenación de los fondos*), cujo sentido mais amplo fica evidente ao discorrer sobre métodos distintos do classificatório, como o de ordem de chegada. O autor trata do código de localização do documento pelo termo *signatura topográfica*, usual em espanhol, apresentando definição genérica representativa dos modos de produzir um código para documentos indicados nas fichas dos mesmos em catálogos.

Na língua portuguesa adotada em Portugal, identificamos o termo *arrumação*, adotado tanto para a ordenação de documentos na estante, quanto para a ordenação de fichas de um catálogo (FARIA; PERICÃO, 2008, p. 73). O termo *cota* é definido como o símbolo usado para indicar o código de localização de documentos em estantes, de modo similar a número de chamada. Por sua vez, as autoras afirmam que, em classificação, especificamente, o termo *número de chamada* é definido a partir da notação clássica composta por número de classificação e elementos de entrada da ficha que o representa (em geral, o nome do autor) (p. 210 e 524).

Em francês, *cote* é o termo adotado para o código que individualiza o documento em uma coleção, como definem Calenge (2009, p. 193) e Boulogne (2005, p. 67). Podemos dizer que *cote* configura-se como elemento da terminologia específica sobre o tema em francês, distinguindo-se de *code*, usado para código no sentido ordinário da língua. O termo técnico francês *cote* é usual na apresentação dos resultados de buscas em bases de dados da França, do mesmo modo que o termo *cota* é adotado em bases de dados de Portugal, para a indicação dos códigos de localização dos documentos representados nos registros.

Em inglês, o termo *shelf mark* indica historicamente a marca ou código escrito ou fixado em um manuscrito ou livro impresso, indicando sua localização em uma biblioteca específica. O termo foi precursor do número de chamada (*call number*),

definido como um código único impresso em uma etiqueta fixada do lado de fora de um item em uma coleção de biblioteca, também mostrado no registro bibliográfico que representa o item no catálogo da biblioteca, para identificar a cópia específica da obra e dar sua localização relativa na estante. Em geral é usado em bibliotecas, segundo estrutura descrita anteriormente, que tem como primeiro elemento de ordenação um número de classificação (REITZ, 2013).

Tanto os termos franceses apresentados anteriormente, relativos aos processos, instrumentos e produtos da ordenação de documentos, quanto os termos adotados em outros idiomas para distinguir ordenação de documentos e classificação bibliográfica, assim como código de localização e número de chamada, demonstram a ausência de um sistema conceitual próprio e de uma terminologia estabelecida sobre o tema no Brasil. No entanto, no livro “A ordem dos livros na biblioteca: uma abordagem preliminar ao sistema de localização fixa”, publicado no Brasil recentemente, o termo ordenação de documentos é adotado e as autoras reconhecem que há negligência no estudo e sistematização da atividade, haja vista tratar-se de recurso para acesso à informação (PINHEIRO, 2007, p. 19). Outro diferencial brasileiro é a literatura sobre os fundamentos das linguagens documentárias, na qual os sistemas tradicionais de classificação bibliográfica são problematizados por meio de aspectos estruturais e funcionais.

Deste modo, se ordenar documentos é propor um arranjo aos mesmos em um certo espaço, a chave da ordenação é o método adotado, o qual pode ser sequencial ou classificatório. A ordenação classificatória, sendo a mais explorada na literatura, apresenta também mais questões para o estudo do tema da ordenação.

A **classificação** em sentido geral implica o processo intelectual pelo qual reconhecemos coisas por apresentarem semelhanças entre si, formando um conjunto de coisas em relação a outro. A **classificação bibliográfica**, por sua vez, é operação intelectual que envolve a identificação de características de documentos, unindo-se aqueles que apresentam características semelhantes e separando-se aqueles com características diferentes. Para o agrupamento de documentos segundo o método classificatório, atributos dos mesmos em uma coleção são adotados, como: tipologia documental (livro, periódico, bibliografia, dicionário, enciclopédia, patente), tipo de conteúdo (gênero literário, biografia), origem da literatura (espanhola, inglesa), público (crianças, mulheres), assunto, outros.

Como podemos ver, classificar documentos não se refere unicamente ao atributo assunto porque não se trata de classificar os conhecimentos dos documentos, mas de ordenar documentos quanto às diversas características pertinentes ao objetivo de comunicação da informação. No entanto, a classificação bibliográfica foi desenvolvida como pertencente aos estudos da chamada Representação Temática. Nessa perspectiva, a compreensão de que a classificação bibliográfica opera unicamente com o atributo assunto foi se sedimentando com o tempo na literatura especializada, no ensino e no pensamento do campo. A classificação é operação lógica de identificação de atributos dos documentos como um todo no conjunto da coleção, e não apenas quanto ao seu conteúdo temático.

Vale retomar a observação de Sayers (1918, cap. V) há um século: analisando a CDD, atentou para o fato de os documentos não tratarem apenas de assuntos específicos, já que muitos, como enciclopédias e jornais, contemplam vários assuntos, e outros, como os livros literários, possuem na forma o modo como os assuntos são apresentados, além de livros que tratam de um assunto segundo uma certa abordagem, como histórica, teórica, outros. Ele afirma ser necessária a distinção entre forma e assunto, propondo uma regra de classificação em que primeiro busca-se pelo assunto, depois pela forma, exceto nos casos em que a forma é primordial. Alguns anos antes, o autor (SAYERS, 1915, p. 43) fez a seguinte sistematização para os casos de uso da forma na classificação bibliográfica: obras compostas de vários assuntos, obras nas quais a forma predomina sobre o assunto e obras em que assuntos específicos são tratados segundo pontos de vista particulares.

Para realizar o processo de classificação bibliográfica, faz-se uso dos **sistemas de classificação bibliográfica** que são, dentre as linguagens documentárias, aqueles que se caracterizam, em termos de sua estrutura, por uma hierarquia, e em termos de sua função, pela vocação para a ordenação de documentos.

Como dissemos, alguns sistemas de classificação bibliográfica são mais amplamente utilizados. Esses sistemas contam, normalmente, com um esquema geral de classificação por grandes temas ou disciplinas, tabelas auxiliares (que se referem a aspectos comuns que podem ser utilizados em qualquer classe do esquema, como forma, tempo, espaço e ponto de vista) e um índice alfabético que facilita o acesso às diversas perspectivas sob as quais um termo pode ser abordado (por exemplo: criança em psicologia, em educação, em antropologia, em religião), assim como, à classe ou subclasse da estrutura hierárquica mais adequada aos objetivos da classificação.

Considerando as vantagens apresentadas no uso dos sistemas de classificação bibliográfica para a ordenação de documentos, relativas aos parâmetros cognitivos que oferecem para uma navegação entre documentos, ao uso grandemente disseminado da codificação alfanumérica e à familiaridade dos públicos quanto à sua estruturação, a exploração de suas características e de suas formas de uso permite ampliar a compreensão sobre esses sistemas, como segue.

Uma dessas características é a centralidade do livro. Sistemas de classificação bibliográfica foram desenvolvidos para dar conta em especial de documentos da tipologia livro, como literatura ficcional, entre outras, e monografias, além de obras de referência (enciclopédias, dicionários) e periódicos. A adequação desses sistemas está relacionada às tipologias documentais recorrentes em bibliotecas no decorrer dos tempos. Para outras tipologias como documentos audiovisuais, relatórios, patentes, entre outros, modos de ordenação e codificação particularizados são considerados mais adequados. Embora menos usual, o termo sistema de classificação documentária é adotado desde há um século para caracterizar a CDU, pois a mesma seria aplicada à ordenação de diversos tipos de documentos, possibilitando a ordenação de repertórios bibliográficos e catálogos; obras em bibliotecas; notas, observações, extratos de documentos diversos destinados ao estudo e trabalhos pessoais; sumários de periódicos; documentos gráficos, ilustrações e fotografias; patentes; catálogos industriais, entre outras aplicações documentárias (CLASSIFICAÇÃO..., 1907). Assim também, Salvan (1962, p. 10) trata da CDU como uma classificação documentária por sua origem, ao mesmo tempo, catalográfica e bibliográfica, voltada a todo tipo de documento. A pretensão era a de que a CDU atendesse à classificação de documentos de qualquer tipo e em qualquer suporte, além de seus metadados. Por outro lado, o termo ‘bibliográfico’ não se refere, em sua origem, exclusivamente a documentos do tipo livro, mas essa relação é usual no campo, evidenciando a vocação dos sistemas tradicionais para livros e monografias em geral.

Quando se trata dos sistemas tradicionais de classificação bibliográfica, entram em questão também as implicações da ênfase à normatividade que ainda hoje caracteriza a mentalidade do campo. O problema se coloca por conta das normas como instância privilegiada de legitimação e compreensão dos procedimentos. A despeito desse pensamento predominante, deve-se buscar pelos fundamentos dos

quais partiram os instrumentos documentários para que os mesmos possam ser utilizados de modo produtivo.

A classificação bibliográfica foi adotada para a ordenação dos documentos por assuntos, tipologias documentais e outros. Posteriormente, a **indexação** passou a ser também adotada por possibilitar, por excelência, a recuperação dos assuntos dos documentos. No entanto, é preciso tratar da classificação bibliográfica e da indexação como duas operações básicas e distintas da organização da informação, como segue.

Como a classificação bibliográfica é usada com o objetivo de realizar a ordenação de documentos, a representação realizada é genérica, pois é baseada no documento como um todo. O produto deste tipo de representação é uma notação classificatória baseada em uma representação linguística referente a uma classe.

Na indexação, a representação realizada é mais específica, pois o objetivo é explorar a diversidade intelectual do documento. O produto da representação, os descritores (ou termos descritores), funcionam como pontos de acesso temáticos que remetem aos registros da base de dados, os quais indicam documentos (a exceção das bases de dados cadastrais que remetem a registros, como dissemos). Diferentemente dos sistemas de classificação bibliográfica, a estrutura classificatória dos tesauros, idealizados para fins de indexação, é apenas uma referência para a organização global do instrumento. O foco na indexação é o conceito, ou uma combinação dos conceitos, não a classe onde ele se localiza. A posição do conceito na estrutura classificatória do tesauro faz com que ele seja marcado pelas características da classe onde se insere mas, diferente da classificação bibliográfica, a classe é relativizada porque o foco desloca-se da estrutura para o próprio conceito representado pelo descritor.

Já na indexação com uso de listas de cabeçalhos de assunto, é comum que cada unidade seja combinada a uma notação classificatória, de modo que não haja duplicação de esforços entre indexação e classificação. O problema é que, além das restrições linguísticas e pragmáticas apresentadas pelos sistemas classificatórios usuais, listas como a *Library of Congress Subject Headings* ou a *Sears List*, segundo Lara (2002, p. 9), são instrumentos que não permitem o controle efetivo do vocabulário porque não dispõem de uma base estrutural de sustentação, o que se agrava quando seus cabeçalhos são traduzidos pelas instituições que os utilizam e registrados em catálogos de rubricas de assunto.

A ordenação de documentos distingue-se da representação descritiva e da representação temática, pois estas operam a partir da exploração dos vários elementos informacionais possíveis e desejáveis para controle e busca, enquanto aquela tem como objeto de representação atributos da unidade documental em sua globalidade. Os aspectos fundamentais da ordenação de documentos são a materialidade dos documentos e a espacialidade relativa ao lugar dos documentos em uma coleção junto a outros que a compõem. Deste modo, a ordenação de documentos refere-se ao trato dos documentos no que tange a um atributo considerado central, enquanto a representação descritiva e a temática operam, respectivamente, com os diversos elementos de identificação e temáticos dos documentos.

A atividade de ordenação em sentido amplo refere-se a um arranjo de documentos ou de metadados de documentos que permite navegação e acesso. Ainda que se possa incluir um mesmo registro em diversas categorias no caso dos metadados dos documentos, a atividade de ordenação nunca oferece o caráter multidimensional de representação que caracteriza as bases de dados, por meio das quais são possíveis várias entradas ou pontos de acesso aos registros dos documentos. A atividade de ordenação de documentos ou de metadados de documentos possui como matriz ordenadora uma estrutura hierárquica ou unidimensional (como desenvolvemos no próximo tópico) ou uma sequência (numérica ou alfabética) que determina o arranjo básico. Nas bases de dados, por meio da representação descritiva e temática, há mais possibilidades de escolhas e pode-se fazer cruzamentos enquanto, a partir das listas ordenadas sequencial ou hierarquicamente, o caminho cognitivo a percorrer é sempre linear.

A ação classificatória é operação adotada em todos os processos de organização da informação – catalogação, indexação, elaboração de resumos e ordenação de documentos e de metadados (sequencial não contínua e sistemática) – , assim como, nos processos subsequentes de busca em bases de dados ou em acervos ordenados. Decorrência da indistinção sobre o papel ocupado pela classificação em cada processo de organização da informação é a cada vez maior equivalência entre **classificação** e **organização da informação**. Se, em meados do século XX, estudos de classificação eram representativos do que hoje compreendemos como organização da informação, uma vez que estava ainda em desenvolvimento um aprofundamento sobre cada um dos seus processos específicos, faltaria, na atualidade, reconhecer os avanços realizados. Os estudos de organização

da informação, assimilados a estudos de classificação, imprimem uma simplificação sobre a compreensão dos seus processos, uma vez que estes são reduzidos unicamente à atividade de identificação de um atributo central do documento, obscurecendo a especificidade de cada atividade, assim como sua função, métodos e problemas particulares. Maniez (2002, p. 231) fala do uso do termo inglês *classification* por Lancaster, quando este se refere à indexação como processo de classificação, exemplificando com a indexação de um documento pelos termos 'escola primária', 'televisão' e 'habilidade de leitura'. Para Lancaster, o exemplo demonstra a atribuição deste documento às três classes identificadas por estes termos, e não, em especial, ao conceito que representam, como tratamos anteriormente. Para Maniez, é preciso distinguir os usos específicos do termo classificação em Ciência da Informação.

5 Fundamentos linguísticos da ordenação de documentos

O desenvolvimento teórico e metodológico das linguagens documentárias, visando subsidiar sua construção e atualização, conta com aportes da Linguística e da Terminologia como modo de promover capacidade linguística de comunicação segundo terminologia adotada pelos públicos. Podemos dizer que Gardin e outros inauguraram uma fundamentação linguística dos processos documentários na década de 1960, embora ensaios já tivessem sido aventados anteriormente.

A definição usual de linguagens documentárias é a de instrumentos adotados para a representação dos conteúdos dos documentos, tendo como fim a recuperação dos mesmos em bases de dados. Para sua funcionalidade efetiva são elaboradas ou adaptadas para uso em cada contexto documentário. Vários tipos de linguagens documentárias foram construídos e utilizados no decorrer de seu desenvolvimento, como os sistemas de classificação bibliográfica, as listas de cabeçalhos de assunto e os tesauros. No entanto, algumas características intrínsecas dos sistemas de classificação bibliográfica, no contexto das linguagens documentárias, e outras características específicas de sistemas mais conhecidos, devem ser evidenciadas.

Uma linguagem documentária é constituída por termos e pelas relações entre eles. Os três tipos de relações que a caracterizam são (CINTRA et al., 2002, p. 45-46):

- relações hierárquicas, cujos termos superordenados e subordinados podem ser do tipo gênero/espécie ou do tipo todo/parte;
- relações de equivalência, nas quais, a um termo preferido corresponde um ou vários não preferidos, sendo que todos funcionam como chave de busca pelo usuário; são as relações que realizam o controle do vocabulário, ou seja, controle da ambiguidade da linguagem natural; e
- relações associativas, aquelas que apresentam algum interesse para busca e que não são as explicitadas anteriormente, como causa e efeito, processo e instrumento etc.
- Sob outro ponto de vista, Gardin (citado por CINTRA, 2002, p. 35-36) trata dos três elementos básicos da linguagem documentária:
- léxico: lista de descritores, devidamente filtrados e depurados;
- rede paradigmática: composta por relações essenciais e, geralmente estáveis, entre descritores, cuja rede lógico-semântica, corresponde à organização dos descritores numa forma que poderia ser chamada de classificação; e
- rede sintagmática: composta pelas relações contingentes entre os termos descritores, relações que são válidas no contexto particular em que são usadas; os sintagmas são as expressões construídas por regras sintáticas cuja coordenação entre termos constitui um tema.

A rede paradigmática representa as unidades que compõem a linguagem documentária quanto às suas categorias (a linguagem documentária propriamente dita), enquanto a rede sintagmática representa o uso destas unidades em combinações pertinentes realizadas na indexação e na busca (a linguagem documentária no uso).

Membro do grupo de pesquisa coordenado por Gardin, Coyaud (1966, p. 22-23) trata de dois tipos de relações entre os termos em uma linguagem documentária:

- relações analíticas: conjunto das relações indicadas de forma explícita entre os termos do léxico para a análise documentária; constituem a parte *a priori*, inata.
- relações sintéticas: aquelas estabelecidas entre os termos do léxico no momento da análise documentária; as ligações particulares que elas

> formam entre os termos do léxico, na indexação dos documentos, não são previstas em sua individualidade, constituindo-se *a posteriori*.

Coyaud indica a organização paradigmática de uma linguagem documentária como aquela fundada em relações analíticas, do que inferirmos que a organização sintagmática seria aquela baseada nas relações sintéticas. Afirma que a diferença entre as relações analíticas e sintéticas não é de natureza, mas de emprego e exemplifica a relação analítica com a relação de inclusão, geralmente presente nas classificações hierárquicas. Segundo ele, Gardin (1964), mas também Grolier (1962), atribuem as relações analíticas ao plano do pensamento e as relações sintéticas ao plano do discurso. Assim, as relações analíticas são aquelas estáveis em uma linguagem documentária, no sentido de que são seus elementos constituintes, e não resultado das combinações realizadas quando do seu uso, caso das relações sintéticas. É preciso observar que, para Gardin, as relações analíticas constituintes da rede paradigmática são sempre relações hierárquicas e as únicas estáveis em função de que seu estudo baseia-se apenas nas linguagens classificatórias, como tratamos a seguir.

Gardin (1966), em estudo sobre as classificações, descarta a observação de codificações e outros aspectos formais (sistema decimal, símbolos alfanuméricos), assumindo que é por meio de seus traços estruturais que parece razoável definir os principais tipos de classificação. É neste contexto que Gardin trata das relações analíticas, as quais definiu como cada relação que une o termo à classe da qual ele faz parte em uma organização lexical, neste caso, uma classificação. Chama de dimensões de uma classificação a natureza das relações analíticas que a constituem, propondo, deste modo, dois tipos de classificações: a unidimensional e a pluridimensional. A primeira apresenta uma relação analítica apenas, enquanto a organização pluridimensional é identificada após a separação das diversas categorias de relações analíticas subjacentes a uma classificação. Coyaud cita Gardin (1962) que, também no texto de 1966, fala em univocidade e multivocidade de uma classificação referindo-se a sua dimensionalidade: uma classificação é unívoca quando um termo aparece em um único lugar e multívoca quando um mesmo termo é alocado em classes diversas, por meio da mesma relação analítica ou de outras, implicando a repetição de termos.

Ao se referir aos sistemas tradicionais de classificação bibliográfica, no entanto, Gardin (1966) propõe distinguir entre uma unidimensionalidade real e uma

unidimensionalidade aparente. A unidimensionalidade real é aquela, por exemplo, das taxonomias, típicas das ciências naturais, em que a relação constitutiva é efetivamente única, em todos os níveis e para todos os grupos da classificação. No entanto, a unicidade de dimensão pode ser aparente. Afirma que, sob a qualificação de hierárquico, frequentemente atribuída a esta relação única, identifica-se tipos diferentes de relação, como qualificação, localização, instrumentalidade etc., que não tem nada em comum com uma relação de inclusão. Para Gardin, enquanto os sistemas taxonômicos apresentam uma unidimensionalidade real, em geral os sistemas de classificação bibliográfica deflagram uma unidimensionalidade aparente, a não ser em algumas de suas partes, como nas nomenclaturas de espécies naturais, de corpos químicos etc. Desse modo, outros tipos de relações ficam anônimas e, por essa razão, são confundidas com a relação hierárquica estrita. Por sua vez, Coyaud (1966, p. 35-376) explora a ideia demonstrando a diversidade não explicitada de relações analíticas da CDU, na qual são amalgamadas diversas relações em uma só relação de inclusão.

Posteriormente, Lara (2002) analisou a CDD e a CDU e realizou comparação entre os dois sistemas, reforçando a análise de Gardin e Coyaud. Afirma que ambos apresentam restrições por incluir, numa única hierarquia, diferentes tipos de relações entre suas unidades. As relações hierárquicas podem ser observadas segundo a natureza da relação (sendo a de gênero/espécie a única relação hierárquica genuína), ou segundo a forma de apresentação (quando tanto a relação gênero/espécie quanto a de todo/parte são apresentadas sob a forma de árvore). Lara (p. 6-7) afirma que o mundo não se organiza de maneira hierárquica (ou seja, de modo unidimensional) e vários tipos de relações podem ocorrer entre os conceitos, como as relações denominadas associativas não-hierárquicas, do tipo causa-efeito, produtor-produto, ação-objeto da ação, agente-instrumento, atividade-agente etc. Para ela (p. 4), sistemas como a CDD e a CDU são mais aptos a prover uma base para a formação de estoques e para a reprodução do conhecimento que para promover a circulação da informação.

Quanto à característica unidimensional dos sistemas hierárquicos explicitada por Gardin, importa distinguir as restrições inerentes aos sistemas de classificação bibliográfica e aquelas específicas dos sistemas tradicionais de classificação bibliográfica, como a CDD e CDU. Os sistemas de classificação bibliográfica são sistemas intrinsecamente hierárquicos, portanto, unidimensionais. A tentativa de

contemplar outras dimensões, por meio da inserção de relações diversas entre os termos, descaracteriza estes sistemas tornando-os unidimensionais aparentes, portanto, comprometidos em sua coerência e funcionalidade. Os sistemas de classificação bibliográfica, assim como, as listas de cabeçalhos de assunto, não são consideradas linguagens documentárias por excelência, caso dos tesauros. O que está em questão é a relação entre a vocação de cada linguagem documentária, o modo como é construída e aquilo para o que é usada. Como exemplo, temos as primeiras propostas de taxonomias corporativas que, rapidamente, incorporaram outros tipos de relações entre os termos para além das relações hierárquicas, visando funcionalidade como instrumento de recuperação da informação.

Neste sentido, a atualidade da contribuição de Gardin, entre outras, está na constatação de que a escolha das relações entre os termos constituintes de uma linguagem documentária manifesta a originalidade de seu ponto de vista estrutural e, consequentemente, seu lugar na tipologia de linguagens em geral. Gardin (1966) avança na identificação de tipologias de classificações pluridimensionais, tratando de uma distinção empírica entre duas categorias de relações analíticas:

- semânticas: referem-se à ordem dos termos (ou seja, das noções que eles designam) segundo definições correntemente admitidas por uma comunidade e consideradas características inerentes das entidades visadas, ou seja, válidas sempre e em qualquer contexto; a classificação pode ser considerada como a imagem dessa ordem (provisória, local), em que a significação primitiva dos termos determina seu lugar na organização geral; são relações essenciais.
- sintáticas: não tratam da essência das entidades, mas de sua função particular em um campo de observação determinado; propriedades consideradas para agrupar os termos são contingentes, ou seja, são válidas somente em um contexto restrito, aquela em que a classificação se desenvolve; são relações funcionais.

Mas Gardin observa que as classificações documentárias são baseadas em relações tanto essenciais quanto funcionais, configurando classificações mistas, ou analítico-sintéticas, as quais variam, desde a ordem estritamente semântica das taxonomias mais puras até a ordem exclusivamente sintática, como as classificações facetadas. Neste último caso, um termo pode constar em qualquer lugar da classificação segundo as diferentes classes funcionais às quais se observa que ele

pertence. Como exemplo, em uma classificação relativa à indústria farmacêutica, o termo iodo pode compor classes que se referem à matéria-prima, ao produto final, ao reagente etc., levando a que o termo se repita na estrutura. A repetição dos termos sob diferentes classes, em virtude de suas diversas dimensões, é típico dos sistemas de classificação facetada. A estrutura hierárquica inclui relações de outro tipo, quais sejam, as de nível sintático pois, para as classes A, B e C, as relações são estabelecidas porque se identifica que há termos que são membros de A, membros de B e membros de C. Tem-se, assim, uma classificação multívoca, já que os termos se repetem em diversos lugares da mesma.

A ideia de faceta, proposta por Ranganathan, inovou ao isolar as características dos termos e organizá-las na estrutura segundo os pontos de vista de interesse. A teoria da classificação facetada, como ficou conhecida, foi uma das bases para o desenvolvimento de linguagens documentárias vocacionadas para a recuperação da informação, como os tesauros.

A aproximação das classificações a uma linguagem propriamente dita, como desenvolveu Gardin, possibilitou uma compreensão do eixo paradigmático e do eixo sintagmático das linguagens documentárias. Este referencial teórico da linguagem permitiu distinguir linguagens documentárias pré-coordenadas, como as classificações facetadas e outras, em que os termos estão combinados previamente, compondo o sistema, ou são minimamente combinados no momento da representação dos documentos, e linguagens documentárias pós-coordenadas, como os tesauros, nas quais a combinação de termos se dá na representação de documentos e busca no sistema. Também é possível observar a distinção entre linguagens documentárias de classificação e linguagens documentárias de indexação, as primeiras voltadas para o documento como um todo, em uma organização linear e unidimensional, e as segundas para seus conteúdos temáticos, em perspectiva multidimensional.

A compreensão linguística da organização da informação não se dá fora de uma perspectiva pragmática, de tal modo que a produção e o uso dos sistemas tradicionais de classificação bibliográfica devem privilegiar características das coleções de documentos e especificidades do público-alvo. Frente à tarefa de fornecer cientificidade a algumas concepções do senso comum, identificamos três questões relacionadas, que comentamos a seguir.

A primeira delas refere-se à caracterização usual dos sistemas tradicionais de classificação bibliográfica como **enciclopédicos**, embora refiram-se de fato “à pretensão universalizante de abarcar, num só sistema, todo o universo do conhecimento” (LARA, 2001). A universalização, no entanto, não ocorre porque, segundo esta autora, em uma abordagem semiótica, o que é registrado “é sempre parcial frente à infinita possibilidade de interpretação enciclopédica”. Lara (1999) debruçou-se sobre o conceito de enciclopédia demonstrando não apenas sua complexidade, mas sua pertinência para uma compreensão da área, em contraposição à ideia de dicionário. Enquanto o dicionário apresenta o léxico de um idioma e suas definições em uma apresentação alfabética, a enciclopédia faz-se pela noção de encadeamento dispondo ao leitor verbetes relacionados que possibilitam diversos percursos de leitura. Posteriormente, Olga Pombo explorou o conceito de enciclopédia como aquele que fornece à área categorias de pensamento que a explicam e sustentam (POMBO, 2010).

Uma segunda questão refere-se a que a **matriz disciplinar** presente nos sistemas tradicionais de classificação bibliográfica não é representativa do mundo contemporâneo, o que dificulta seu uso e atualização. Assim, se forçar um sistema hierárquico para que seja multidimensional não se apresenta viável, a tentativa de englobar nele todo o conhecimento relativo a um período de tempo bastante antigo, evidencia visão de mundo homogênea e estável e, portanto, anacrônica.

A terceira questão é que, na cultura biblioteconômica, a informação foi continuamente considerada um dado, ou seja, a informação existiria *a priori* e a representação seria o esforço de resgatá-la do documento. Esta perspectiva levou à ideia de que na representação há uma relação de fidedignidade entre o conteúdo de um documento e a expressão utilizada para ‘representá-lo’. Como decorrência, com base em Lara (2001), o produto da classificação seria uma **reprodução**.

Quanto às três questões, observamos, respectivamente, que persiste a ideia de sistemas pretensamente universais, adotados na ausência da percepção do significado da datação histórica de uma determinada abordagem do conhecimento, e que partem de pontos fixos de enunciação a serem mecanicamente equiparados aos conteúdos dos documentos, cuja interpretação também seria única.

Para Lara (2001), a informação é sempre uma construção, porque as generalizações realizadas na atividade documentária são um produto cultural, fruto da experiência com a realidade, portanto, não são neutras, nem possuem um fim em si

mesmas. A constituição de sistemas de informação é uma operação carregada de intencionalidade que persegue objetivos específicos. As necessidades pragmáticas determinam os recortes, ou seja, as segmentações dos conteúdos, de forma a viabilizar um fluxo informativo efetivo. Consequentemente, um mesmo universo documentário pode ser organizado e representado de diversas maneiras, pautando-se nos distintos propósitos a que se destinam.

Se a linguagem é a principal expressão das diferentes formas de organização cultural, justifica-se “procurar nas teorias da linguagem o apoio para a construção das linguagens documentárias, bem como as referências para alcançar formas de mediação mais adesivas aos sistemas informacionais” (MOREIRA; LARA, 2012).

Desse modo, sob o ponto de vista linguístico, a ordenação classificatória baseia-se em uma estrutura unidimensional, a qual apresenta características próprias que nos permite avaliar com rigor os diversos sistemas de classificação bibliográfica. As implicações de sua aparente unidimensionalidade passam pela constatação de estes sistemas não serem, de fato, nem unidimensionais, possuindo apenas um tipo de relação de analítica, nem pluridimensionais, apresentando multivocidade a partir de repetições coerentes, organizadas. É preciso distinguir aspectos essenciais e contingenciais, os primeiros sempre necessários à percepção destes últimos. A análise de Gardin permite identificar que os problemas dos sistemas tradicionais de classificação bibliográfica não são intrínsecos a este tipo de sistema. O fato de alguns deles não serem efetivamente unidimensionais é uma contingência.

6 Considerações finais

Processos e instrumentos baseados em sequências e/ou em hierarquias são pertinentes à ordenação de documentos e de metadados dos documentos pois, trata-se, em todos os casos, de alocação no espaço. Para a atividade de representação de documentos visando recuperação de seus conteúdos, a multidimensionalidade, característica dos tesauros adotados para a operação da indexação, deve ser a referência.

Não há um único modo de ordenação de documentos, embora seja produtivo compartilhar soluções. Faz-se necessária a produção de arranjos que considerem usuários, documentos e objetivos institucionais. Como em todas as outras operações documentárias, a orientação é sempre pragmática, ou seja, baseada em sua possível

utilidade. O desafio colocado é o do refinamento de metodologias baseadas em parâmetros que permitam a elaboração de categorias pertinentes à atividade.

A escolha de categorias é sempre local e provisória. Paradoxalmente, o trabalho documentário necessita de estruturas de representação relativamente estáveis, em função do alto custo da recodificação de documentos e da exigência de compreensão dos novos modos de ordenação pelos usuários.

Em ambientes tradicionais ou eletrônicos, nenhuma das operações descritas para a atividade de ordenação é de fato mecânica ou de menor exigência intelectual. A ordenação de documentos deve ter seu papel revisto no contexto do conjunto das ações de mediação realizadas nos diversos serviços de informação. Elaborar os processos de organização da informação em sua essencialidade fornece base sólida de sustentação para operar com as mudanças que ocorrem no tempo.

À falta de olhares suficientemente problematizadores sobre a ordenação bibliográfica de documentos no ensino e na pesquisa, realizamos duas observações. A primeira, sobre a aparente contradição entre a ênfase aos usuários e o ensino inquestionado de modelo único como solução para a ordenação de documentos; de fato, os estudos sobre usuários nem sempre consideram a perspectiva mediadora dos processos de organização da informação que permite o vínculo com um público. A segunda observação é a de que há produção de literatura fundamentadora e operacionalizadora do processo de ordenação dos documentos, desde os seculares estudos sobre bibliografias e bibliotecas, até os aprofundamentos linguísticos dos instrumentos classificatórios propostos por Gardin na metade do século passado e seus desenvolvimentos, assim como, a proposta contemporânea francesa, cuja concepção e aplicações merecem ser conhecidas.

Referências

BOULOGNE, Arlette (Coord.) (2005). **Vocabulaire de la documentation.** Paris: ADBS Éditions. (Sciences et techniques de l'Information). 334 p. Também disponível no endereço: <http://www.adbs.fr/vocabulaire-de-la-documentation-41820.htm?RH=OUTILS_VOC&RF=OUTILS_VOC>. Acesso em: 1 ago. 2015.

BRIET, Suzanne (1951). **Qu'est-ce que la documentation?** Paris: Édit - Éditions Documentaires Industriales et Techniques. 48 p. Disponível em: <http://martinetl.free.fr/suzannebriet/questcequeladocumentation/briet.pdf>. Acesso em: 1 ago. 2015.

CACALY, Serge et al (2008). **Dictionnaire de l'information.** 3. ed. Paris: Armand Colin. 296 p.

CALENGE, Bertrand (Coord.) (2010). **Mettre en oeuvre um plan de classement.** Villeurbanne: Presses de l'ENSSIB. (La Boîte à outils, 18). 199 p. p. 191-196: Glossário. Publicado originalmente em 2009.

CARRIÓN GÚTIEZ, Manuel (reimpr. 2002). **Manual de bibliotecas.** 2. ed. de 1993. Madrid: Fundación Germán Sánchez Ruipérez. 766 p. p. 391-402.

CINTRA, Anna Maria Marques et al. (2002). **Para entender as linguagens documentárias.** 2. ed. São Paulo: Polis.

CLASSIFICAÇÃO bibliográfica decimal (1907). **Bulletin de l'Institut International de Bibliographie**, Bruxelles, ano 12, n. 68-70, não paginado.

COMBOT, Philippe (2011). **Initiations aux techniques documentaires**. Disponível em: <http://www.youscribe.com/catalogue/tous/education/cours/initiation-aux-techniques-documentaires-385019> e em <http://combot.univ-tln.fr/lea/index.html>. Acesso em: 1 ago. 2015.

COYAUD, Maurice (1966). **Introduction a l'étude des languages documentaires.** Paris: C. Klincksieck. 148 p. (T. A. Documents).

FARIA, Maria Isabel Ribeiro de; PERICÃO, Maria da Graça (2008). **Dicionário do livro**: da escrita ao livro eletrônico. São Paulo: EDUSP. 768 p.

GARDIN, Jean-Claude (1966). Elements d'un modele pour la description des lexiques documentaires. **Bulletin des Bibliotèques de France***,* n. 5, p. 171-182. Disponível em: <http://bbf.enssib.fr/consulter/bbf-1966-05-0171-001>. Acesso em: 1 ago. 2015.

HUDON, Michèle (2009). Le traitment du document. In: SALAÜN, Jean-Michel; ARSENAULT, Clément (Coords.). **Introduction aux sciences de l'information.** Montréal: Les Presses de la Université de Montréal. (Paramètres). 235 p. p. 53-100.

JOUGUELET, Suzanne (1997). Classement. In: CACALY, Serge (Coord.). **Dictionnaire encyclopédique de l'information et de la documentation.** Paris : Nathan. 635 p. p. 135-136.

LARA, Marilda Lopes Ginez de (1999). **Representação e linguagens documentárias**: bases teórico-metodológicas. São Paulo. Tese (Doutorado) – Escola de Comunicações e Artes/USP.

LARA, Marilda Lopes Ginez de (2001). O unicórnio (o rinoceronte, o ornitorrinco...), a análise documentária e a linguagem documentária. **DataGramaZero**: Revista de Ciência da Informação, Rio de Janeiro, v. 2, n. 6.

LARA, Marilda Lopes Ginez de (2002). **Dos sistemas de classificação bibliográfica às *search engines* (I).** São Paulo: Associação Paulista de Bibliotecários. (Ensaios APB, 90).

MANIEZ, Jacques (2002). **Actualité des langages documentaires**: les fondements thèoriques de la recherche d'information. Paris: ADBS. (Collection Sciences de l'Information, Série Recherche e documents).

MENESES, Ulpiano T. Bezerra de (1994). Do teatro da memória ao laboratorio da história: a exposição museológica e o conhecimento histórico. **Anais do Museu Paulista**, v. 2, n. 1, p. 9-42. Disponível em: <http://www.scielo.br/pdf/anaismp/v2n1/a02v2n1.pdf>. Acesso em: 1 ago. 2015.

MEYRIAT, Jean (1981). Document, documentation, documentologie. **Schéma et Schématisation**, 2º trimestre, n. 14, p. 51-63.

MOREIRA, Walter; LARA, Marilda Lopes Ginez de (2012). Ontologias, categorias e interoperabilidade semântica. **DataGramaZero**, v. 13, n. 4.

PERROTTI, Edmir; PIERUCCINI, Ivete (2007). Infoeducação: saberes e fazeres da contemporaneidade. In: LARA, M. L. G. de; FUJINO, Asa; NORONHA, Daisy Pires (Orgs.). **Informação e contemporaneidade**: perspectivas. Recife: Néctar. p. 47-98. Disponível em: <http://www3.eca.usp.br/sites/default/files/form/ata/pos/ppgci/publicacoes%20-%20informacaoContemporaniedade%281%29.pdf> . Acesso em: 1 ago. 2015.

PINHEIRO, Ana Virginia Teixeira da Paz (2007). **A ordem dos livros na biblioteca**: uma abordagem preliminar ao sistema de localização fixa. Rio de Janeiro: Interciência, Niterói: Intertexto.

POMBO, Olga (2010). Dispersão e unidade: para uma poética da simpatia. In: LARA, Marilda Lopes Ginez de; SMIT, Johanna Wilhelmina (Orgs.). **Temas de pesquisa em Ciência da Informação no Brasil.** São Paulo: Escola de Comunicações e Artes/USP. p. 29-46. Disponível em: <http://www3.eca.usp.br/sites/default/files/form/ata/pos/ppgci/publicacoes%20-%20temasdepesquisas.pdf>. Acesso em: 1 ago. 2015.

REITZ, Joan (2013). **ODLIS** - Online Dictionary for Library and Information Science. Disponível em: <http://www.abc-clio.com/ODLIS/about.aspx>. Acesso em: 1 ago. 2015.

SAGREDO FERNÁNDEZ, Felix; IZQUIERDO ARROYO, José María (1983). Concepción lógico-lingüística de la Documentación. Madrid: IBERCOM. 440 p.

SALVAN, Paule (1962). **Les classifications**. 2. ed. Paris: Bibliothèque National. Preparação para o Diploma Superior de Bibliotecário. Disponível em: <http://www.enssib.fr/bibliotheque-numerique/documents/48774-les-classifications.pdf>. Acesso em: 1 ago. 2015.

SAYERS, W. C. Berwick (1915). **Canons of classification applied to "The subject" "The expansive", "The decimal" and "The Library of Congress" classifications**: a study in bibliographical classification method. London: Grafton. 173 p.

SAYERS, W. C. Berwick (1918). **An introduction to library classification**: with readings, questions and examination papers. London: Grafton & Co. ; New York City: The H. W. Wilson Co. 172 p.

WEITZEL, Simone da Rocha (2012). Desenvolvimento de coleções: origem dos fundamentos contemporâneos. **TransInformação**, Campinas, v. 24, n. 3, p. 179-190.

A ISKO-Brasil e o fortalecimento de intercâmbios e visibilidade científica em organização do conhecimento[25]

Vera Dodebei
Universidade Federal do Estado do Rio de Janeiro
dodebei@gmail.com

O campo da organização e representação do conhecimento (OC), segundo Dodebei e Guimarães (2013), vem se tornando um desafio à medida que as bases de registro se virtualizam e a informação é lançada na rede de computadores em múltiplos formatos e com múltiplas etiquetas. Se, por um lado, isto garante a diversidade cultural no uso dos recursos, por outro lado, o alto índice de dispersão temática desses recursos de informação vai obrigar os pesquisadores da organização do conhecimento a desenharem, em conjunto com profissionais de campos correlatos, sistemas e programas que auxiliem os processos de indexação e classificação compartilhados. A colaboração entre produtores e usuários do conhecimento no mundo contemporâneo parece ser a melhor estratégia para realizar tarefas altamente dispendiosas como aquelas típicas da área da OC nos exemplos da classificação bibliográfica, da indexação e da elaboração de linguagens documentais ou documentárias.

A OC é um campo que demanda reflexão constante no sentido da criação de sistemas de conceitos desenvolvidos “sob medida” para organizar acervos analógicos e digitais, websites, repositórios, banco de dados com finalidades acadêmicas e comerciais e, nesse sentido, a troca de experiência com pesquisadores de centros de pesquisa internacionais é fundamental para a ampliação da fundamentação teórica-epistemológica da pesquisa técnico-científica.

[25]Mesa Redonda: Visibilidade científica einterlocução internacional em organização do conhecimento. III ISKO-Brasil. Marília, SP. 2015.

A Sociedade Brasileira para a Organização do Conhecimento (ISKO-Brasil), http://isko-brasil.org.br/criada em 30 de outubro de 2007, constitui o capítulo brasileiro da “Internationale Gesellschaft für Wissensorganisation”,ou International Society for Knowledge Organization (ISKO)http://www.isko.org/index.php, registrada na comarca de Frankfurt, Alemanha, em 22 de julho de 1989.A ISKO-Brasil é uma associação de âmbito nacional, com sede jurídica em Marília, SP, onde foi criada, e tendo como sede administrativa a instituição onde pertence o presidente, atualmente, no mandato de 2013-2105, no Rio de Janeiro.

A ISKO-Brasil tem por fim promover a pesquisa, o ensino, o desenvolvimento e a aplicação de métodos para organizaçãodo conhecimento (OC) em geral e em campos específicos nos seguintes aspectos: promover em nível nacional e internacional o intercâmbio de informação e de experiências entre pesquisadores e profissionais na área de OC, realizar reuniões e eventos científicos na área, organizar epromovera visibilidade da pesquisa em publicações na área, colaborar com a International Society for Knowledge Organization na consecução de seus objetivos e promover mecanismos de cooperação científica, bem como a promoção e a inserção internacional de estudos epesquisas realizadas por seus membros.

Nossos objetivos, a curto prazo, vêm sendo fortalecer o vínculo institucional entre as duas associações. Tal vínculo manifestar-se-á durante os próximos anos com a execução das seguintes metas e ações:1) organizar dois eventos nacionais (2015 e 2017) em que estão previstas atividades com convidados internacionais, como palestras, minicursos, debates com grupos de pesquisa. Nesses eventos estão previstas reuniões de consolidação de grupos de pesquisa já em desenvolvimento e estímulo à formação de novos grupos, agregando pesquisadores brasileiros e estrangeiros;2) organizar um evento internacional em 2016, no qual estão previstos minicursos com especialistas internacionais, além de duas conferências magnas: de abertura e de encerramento; e, 3) publicaros anais eletrônicos com as comunicações apresentadas nos eventos nacionais, o que dinamiza os instrumentos de cooperação, por intermédio ao estímulo a publicações em coautoria, intensificando as parcerias em andamento e estimulando novas parcerias.

Em seus quase vinte anos de existência, a ISKO Internacional expandiu-se de forma considerável em todos os continentes, compondoseu quadro de associados representantes dos seguintes países: Alemanha, Argélia, Benim, Brasil, Camarões, Canada, China, Espanha, Estados Unidos, França, Gana, Índia, Inglaterra, Irã, Itália,

Marrocos, Nigéria, Polônia, Singapura, Taiwain e Tunísia.A informação mais recente indica que seu quadro de associados conta com 596 pesquisadores, entre os quais pesquisadores muito experientes e outros, em menor número, de pesquisadores em formação.Vale lembrar que 18.3% desse total são oriundos de países em desenvolvimento.Importa ressaltar também o papel que o Brasil representa no quadro geral:

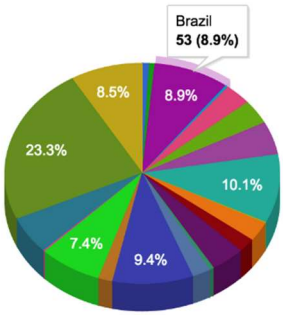


Gráfico 1 – Associados ISKO

Fonte: ISKO. Disponível em http://www.isko.org/index.php. Acesso em: 08.03.2015.

Em menos de dez anos de criação, o quadro de nossos associados compõe quase 9% do quantitativo de associados da Associação internacional, superado apenas por 3 países (Inglaterra, Alemanha e Polônia), o que indica uma forte presença de investigação científica dessa área em nosso país. Além disso, é digna de nota a preocupação com a formação de novos pesquisadores. No quantitativo de sócios, o Brasil conta com aproximadamente 20% de associados que estão cursando a sua formação em mestrado ou doutorado (conforme quadro de associados disponível no website da associação), o que nos indica uma revitalização da área de pesquisa e a certeza de que o campo investigativo continua produtivo.

O capítulo brasileiro promoveu, desde sua criação em 2007, nos anos não coincidentes ao evento internacional, a organização de alguns eventos. No ano de 2009, ainda por conta de ser uma associação incipiente, realizou um pequeno evento inserto no GT2 do ENACIB, conformado segundo a sua origem inicial. Nesse evento, dada a repercussão e a abrangência de participantes, foi constata pela diretoria executiva a possibilidade de alçar voos mais consistentes, quando os associados decidiram organizar o I evento nacional em 2011 em Brasília. Em 2013, a edição do II Congresso nacional teve lugar no Rio de Janeiro.Nesta III edição o Congresso Nacional realiza-se em Marília e a 14ª.edição da conferência internacional será no Brasil, na cidade do rio de Janeiro.

Do ponto de vista da pesquisa e, em virtude de a Associação se configurar pela inserção individual de seus membros, existe forte envolvimento de diversas instituições de pesquisa disseminadas por todo o mundo aos quais seus membros estão vinculado. A título de ilustração, podemos apresentar as diversas instituições envolvidas no Brasil, a partir da filiação de seus associados: Universidade Federal de Bahia, Universidade Federal Fluminense, Universidade Federal de Minas Gerais, Universidade Estadual de Londrina, Universidade de São Paulo, Universidade Federal do Estado do Rio de Janeiro, Universidade Federal de Pernambuco, Universidade de Brasília, EMBRAPA, INMETRO, Universidade Federal do Paraná, IBICT, Universidade Estadual Paulista, Fundação Getúlio Vargas, entre outras.

Um dos indicadores que poderemos utilizar para aferir o impacto provocado pela cooperação entre a Associação internacional e o capítulo brasileiro é o reflexo no número de associados após cada edição dos eventos científicos.Além desse, vimos constatando presença de pesquisadores brasileiros cada vez mais frequente em programas de intercâmbio interinstitucionais representados por Minter e Dinter (Capes) e professores visitantes, embora ainda seja muito tímida a publicação de artigos no periódico internacional KO.Como dois indicadores, podemos verificar a presença de coautorias de pesquisadores brasileiros entre si e a de coautorias entre pesquisadores nacionais e internacionais e, ainda, o estabelecimento do índice de correlações entre pesquisadores e respectivas instituições, no intuito de elencar as redes existentes — e as novas redes que porventura se criem — na área de OC.

Temos como meta a melhoria da qualidade de nossos eventos, no sentido de oferecer aos pesquisadores a oportunidade do contato pessoal com um número maior de grupos de pesquisa para intercâmbios futuros, na medida em que podemos ampliar o convite para a vinda ao Brasil de conferencistas estrangeiros, a publicação da série “Estudos avançados em organização e representação do conhecimento” em formato impresso, além de arquivo aberto, visando a intensificara socializaçãodos resultados de pesquisas e a oportunidade de intercambiar novas tecnologias da informação e comunicação já desenvolvidas e em desenvolvimento no Brasil.

Nesse terceiro congresso, a ISKO-Brasil escolheu discutir desafios e perspectivas que se colocam para a organização do conhecimento frente à diversidade cultural, ao estado da arte da pesquisa brasileira e à interlocução científica no ambiente acadêmico brasileiro e internacional. Os trabalhos selecionados estão já publicados em IPUB e disponíveis no website da ISKO-Brasil, compondo a série:

Estudos avançados em organização e representação do conhecimento: diversidade cultural. v.3, 2015, disponíveis online na página da associação. O planejamento e realização da 14ª. Conferência Internacional da ISKO, no Rio de Janeiro, de 12 a15 de setembro 2016 contará com inúmeros desafiospois será um evento de grande porte, visto que estará aberta submissão de trabalhos aos associados da Isko que hoje congrega 21 capítulos e 596 pesquisadores.

O tema do congresso - *Knowledge organization for a sustainable world: challenges and perspectives for cultural, scientific and technological sharing in a connected society* – coloca o Brasil à frente nas relações do campo da OC, antes restrito às bibliotecas, com as novas tecnologias da comunicação massiva e em rede, e também com as pesquisas em humanidades digitais, memória social e sustentabilidade econômica, o que aponta para novos temas transversais de pesquisa, por exemplo, ecologia da informação, que terão visibilidade com a publicação dos anais impressos e em acesso aberto na língua inglesa, compondo o volume 15 da *Advances in Knowledge Organization*, a ser editado em 2016 pela editora Ergon Verlag, na Alemanha. Oplanejamento e a realização do IV Congresso Brasileiro em *Organização e Representação do Conhecimento* previsto para acontecer em 2017, coloca a Isko-Brasil no calendário de eventos apoiado por agências de fomento e reafirma a classificação de evento bem qualificado no âmbito da avaliação brasileira do ensino de pós-graduação.

Não temos ainda um estudo feito para a visibilidade das cooperações interintitucionais estrangeiras quanto às parcerias de mobilidade estudantil e docente em estágios no exterior e no Brasil e precisamos reforçar o intercâmbio em organização do conhecimento entre grupos de estudo e laboratórios de pesquisas internacionais cujos temas possam fazer avançar as pesquisas na área da organização do conhecimento.

Referências

DODEBEI, Vera; GUIMARÃES, J. A. C. Prefácio. In: *Complexidade e organização do conhecimento: desafios de nosso século.* Rio de Janeiro: ISKO-Brasil; Marília: FUNDEPE, 2013. (e-Book ISBN: 978-85-98176-51-2)

ISKO.http://www.isko.org/index.php

ISKO-Brasil. http://isko-brasil.org.br/

A organização do conhecimento no contexto da pós-modernidade sob a perspectiva da teoria da classificação

Daniel Abraão Pando
Universidade Estadual Paulista
danielabraaopando@gmail.com

Carlos Cândido de Almeida
Universidade Estadual Paulista
carlosalmeida@marilia.unesp.br

Resumo: Aborda as questões relacionadas à temática da organização do conhecimento (OC) no contexto da pós-modernidade sob a perspectiva da teoria da classificação. Tendo como delineamento metodológico uma análise bibliográfica buscou-se registrar o teor dessas questões na literatura nacional e internacional a partir do contraponto com a modernidade. Constata-se que: a) enquanto na concepção moderna tem-se a ideia de que a classificação pode ser um espelho neutro e objetivo de um universo do conhecimento a teoria pós-moderna rejeita tal hipótese deslocando seu foco sobre a práxis social e a linguagem da comunidade; b) enquanto a classificação moderna visava representar o universo do conhecimento a classificação pós-moderna visa fornecer uma ferramenta pragmática para domínios específicos; c) se na modernidade a classificação tinha uma centralidade na OC devido a fisicalidade dos documentos, com o advento das novas tecnologias e um novo espaço de produção de informações digitais os estudos relacionados a classificação parecem sofrer um deslocamento. Aponta-se como tendência para a área que os estudos ligados a Indexação podem se tornar centrais nesse contexto.

Palavras-chave: Organização do conhecimento. Modernidade. Pós-Modernidade. Teoria da Classificação.

Abstract: This paper deals with issues related to Knowledge Organization (KO) in the context of postmodernity from the theory of classification perspective. The methodology herein consisted of bibliographic analyses in order to register these issues contents in national and international literature at the other end of the spectrum of modernity. It is possible to verify that: a) while in the modern conception there is the idea that the classification can be a neutral and objective mirror of a universe of knowledge, the postmodern theory rejects such hypothesis moving its focus on the social praxis and the community language; b) while the modern classification aimed at representing the universe of knowledge, the postmodern classification aims at providing a pragmatic tool for specific domains; c) if in modernity the classification focused on KO due to the inspection of documents, with new technology and a new space to produce digital information, studies related to classification seem to be displaced. We can consider a trend for the area that the studies connected to Indexation can take a very important part in this context.

Keywords: Knowledge Organization. Modernity. Postmodernity. Theory of Classification.

O trabalho na íntegra será publicado na revista Knowledge Organization.

O princípio da garantia semântica e os estudos da linguagem: reflexões conceituais

Roger de Miranda Guedes
Universidade Federal de Minas Gerais
rogerotoni@gmail.com

Maria Aparecida Moura
Universidade Federal de Minas Gerais
mamoura@ufmg.br

Resumo: Analisa aspectos conceituais do princípio da garantia semântica na organização do conhecimento, através dos estudos da linguagem. Para tal análise recorre-se às percepções da linguagem como fenômeno da atividade social, relacionadas às visões de Mikhail Bakhtin e Ludwig Wittgenstein, a partir da concepção pragmática de linguagem dos autores busca-se atualizar as possibilidades de entendimento do princípio da garantia semântica com o intuito de atender as insuficiências de reflexão teórica deste tópico, sobretudo a partir da emergência de fenômenos de informação na Web.

Palavras-chave: Garantia semântica. Estudos da linguagem. Sistemas de Organização do Conhecimento. World Wide Web.

Abstract: It analyzes conceptual aspects of the principle of semantic warranty in the knowledge organization, through language studies. For this analysis refers to the perceptions of language as a phenomenon of social activity related to the views of Mikhail Bakhtin and Ludwig Wittgenstein. From the pragmatic design language of the authors, seek to upgrade the possibilities of understanding the principle of semantic warranty with the order to meet the shortcomings of theoretical reflection of this topic, especially since the emergence of Web information phenomena.

Keywords: Semantic warranty. Studies of language. Knowledge Organization Systems. World Wide Web.

O trabalho na íntegra será publicado na revista Knowledge Organization.

Interoperabilidade e Mapeamentos entre Sistemas de Organização do Conhecimento: Metatesauro-Unified Medical Language System-National Library of Medicine

Julietti de Andrade
Diretora do Serviço de Biblioteca e Documentação Científica e Didática IOT-HCFMUSP
julietti.andrade@gmail.com

Marilda Lopes Ginez de Lara
Universidade de São Paulo
larama@usp.br

Resumo: Esse trabalho apresenta análise sobre a consistência semântica de agrupamento de sinônimos de termo selecionado do Metatesauro - *Unified Medical Language System* da *National Library of Medicine* (UMLS) o qual é resultado de mapeamento entre Sistemas de Organização do Conhecimento (SOC), ou *Knowledge Organization Systems* (KOS), na área da Saúde. Explora as características da equivalência conforme a norma ISO 25964:2:2011 e com base nas categorias: variações semânticas, morfológicas, sintáticas e tipográficas. Destaca a importância do mapeamento e seleção de termos para recuperação da informação e apresentação de resultados de busca em metodologias de trabalhos científicos.

Palavras-chave: Interoperabilidade. Mapeamento entre Sistemas de Organização do Conhecimento. Organização e Recuperação da Informação e do Conhecimento.

Abstract: This work presents analysis of the semantic consistency of term of synonyms grouping selected from Metathesaurus, the Unified Medical Language System of the National Library of Medicine which is the result of mapping between Knowledge Organization Systems (SOC), or Knowledge Organization Systems (KOS) in Health. Explores the characteristics of equivalence, according to ISO 25964: 2:2011 and based on the categories: semantic, morphological, syntactic and typographical variations. It highlights the importance of mapping and selection of terms for information retrieval and presentation of search results in scientific work methodologies.

Keywords: Interoperability. Mapping between Organizing Systems of Knowledge. Organization and Retrieval of Information and Knowledge.

O trabalho na íntegra será publicado na revista Knowledge Organization.

Aspectos metodológicos nas teorias críticas em Organização do Conhecimento

Daniel Martinez-Ávila
Universidade Estadual Paulista
dmartinezavila@marilia.unesp.br

Rafael Aparecido Moron Semidão
Universidade Estadual Paulista
rafaelsemidao@gmail.com

Marcio Ferreira da Silva
Universidade Estadual Paulista
soimers@gmail.com

Resumo: Busca explicitar a configuração metodológica das teorias críticas no domínio da Organização do Conhecimento com o propósito de oferecer categorias de estudo e elementos de cotejo que possam propiciar uma melhor compreensão de tais teorias. Para tanto, descreve as teorias críticas como sendo uma resposta a problemas éticos verificados nos sistemas de classificação universalistas. Estabelece uma caracterização de tipos expressivos de teorias críticas a partir de três grupos sociais não contemplados pelos sistemas universalistas. Apresenta a configuração inferida como sendo uma dinâmica metodológica constituída de três etapas: 1) aporética; 2) de referenciação teórica; e 3) propositiva.
Palavras-Chave: Teorias Críticas. Organização do Conhecimento. Metodologia.

Abstract: This paper aims to reveal the methodological configuration of critical theories in the knowledge organization domain. The goal is to facilitate the study and understanding of these theories. For this, we describe these critical theories as a response to the ethical problems that can be exemplified in the universalist classification systems. We establish a characterization of the expressive types of critical theories based on three social groups that are not traditionally favored by universalist systems. As a result, we present the inferred configuration as a methodological dynamics composed of three stages: 1) aporetics; 2) theoretical framework; and 3) propositive.
Keywords: Critical theories. Knowledge Organization. Methodology.

O trabalho na íntegra será publicado na revista Knowledge Organization.

Propuesta de representación del Tesauro de la UNESCO para la web semántica mediante ISO-25964

Juan-Antonio Pastor-Sánchez
Universidad de Murcia
pastor@um.es

Resumen: Se muestra como se ha aplicado la norma ISO-25964 para representar el tesauro de la UNESCO mediante tecnologías de la web semántica. A partir de los trabajos realizados en el proyecto UNESKOS se ha analizado el uso conjunto de las ontologías SKOS e ISO-THES para representar tesauros conforme al modelo de datos de la norma ISO-25964. El resultado ha sido un conjunto de datos RDF, accesible según los principios Linked Open Data y que hace uso de las ontologías anteriormente mencionadas junto con una serie de propiedades desarrolladas en el contexto del proyecto UNESKOS. Finalmente se apuntan una serie de conclusiones que apuntan, entre otros aspectos, a una revisión de SKOS y a la adopción de tecnologías adecuadas que faciliten el desarrollo de líneas de trabajo enfocadas en la alineación de vocabularios.

Palabras clave: Tesauro de la UNESCO. SKOS. Web Semántica. ISO-25964. ISO-THES. UNESKOS.

Abstract: This paper shows how has been applied ISO-25964 standard to represent the thesaurus UNESCO through semantic web technologies. Based on the work done in the UNESKOS project has been analyzed the joint use of SKOS and ISO-THES ontologies to represent thesauri according to the data model of the ISO-25964 standard. The result has been a set of RDF data, accessible as Linked Open Data principles and making use of the above ontologies along with certain properties developed in the context of UNESKOS project. The conclusions point, among other aspects, a review of SKOS and adoption of appropriate technologies that facilitate the development of future work and research lines focused on the alignment of vocabularies.

Keywords: UNESCO Thesaurus. SKOS. Semantic Web. ISO-25964. ISO-THES. UNESKOS.

Trabalho ganhador do prêmio ISKO Brasil 2015, a ser publicado na íntegra na *revista Brazilian Journal of Information Science*.

Modelagem Metaconceitual: aplicação à Infraestrutura de Qualidade com base na ISO 25964

Catarina Felix dos S. Soares
Instituto Brasileiro de Geografia e Estatística
cathifelix@gmail.com

Vera Dodebei
Universidade Federal do Estado do Rio de Janeiro
dodebei@gmail.com

Resumo: Apresenta proposta metodológica para construção de um sistema de organização do conhecimento (SOC) em Infraestrutura de Qualidade (IQ). O domínio IQ refere-se às atividades de avaliação da conformidade (AC), normalização e metrologia, que no Brasil têm como representante o Instituto Nacional de Metrologia Qualidade e Tecnologia (Inmetro). Utilizou-se vocabulários especializados como fonte conceitual para a construção do tesauro. A pesquisa, de natureza teórico-prática, pautou-se na análise da literatura especializada sobre modelagem conceitual, nas diretrizes de construção de sistemas de organização do conhecimento presentes na norma ISO 25964 (information and documentation and thesauri and interoperability with other vocabularies) e nas teorias da classificação facetada (Ranganathan, 1967), do conceito (Dahlberg, 1978) e da terminologia (Würster, 1981). A integração das teorias e normas técnicas possibilitou a construção de um tesauro metaconceitual, com o qual espera-se aumentar o potencial de utilização da terminologia técnica da área para indexar a literatura, além de oferecer um modelo formal e padronizado para utilização em diferentes SOCs.
Palavras-chave: Modelos conceituais e metaconceituais. Sistemas de Organização do Conhecimento (SOCs). Infraestrutura de Qualidade (IQ).

Abstract: It presents a methodological proposal for building a knowledge organization system (SOC) in Quality Infrastructure (IQ). The IQ domain refers to the conformity assessment activities (AC), standardization and metrology, which in Brazil is represented by the National Institute of Metrology Quality and Technology (Inmetro). Specialized vocabularies was used as conceptual source for thesaurus construction. The research, based on theoretical and practical aspects, considered the literature review on conceptual modeling, knowledge organization systems construction guidelines presented in ISO 25964 (information and documentation and interoperability with other thesauri and vocabularies) and the theories of faceted classification (Ranganathan, 1967) of concept (Dahlberg, 1978) and of terminology (Würster, 1981). The integration of theories with technical standards has enabled the construction of a metaconceptual thesaurus, which is expected to increase the potential use of technical terminology from this specialized area, in order to index literature and to provide a formal and standard model for use in different SOCs.
Keywords: conceptual and metaconceptual models. Knowledge Organization Systems (SOCs). Quality Infrastructure (CI).
Trabalho ganhador do prêmio ISKO Brasil 2015, a ser publicado, na íntegra, na *revista Brazilian Journal of Information Science*.

Tres enfoques de *bias* en organización del conocimiento: *bias* neutro, *bias* negativo y *bias* positivo

Stephanie Colombo
Universidad de la República
stephanie.colombo@fic.edu.uy

Mario Barité
Universidad de la República
mabarite@gmail.com

Resumo: O termo bias é um termo problema na Organização do Conhecimento, principalmente pela falta de uma definição clara. Na literatura da área pode-se encontrar três aspectos, bias com carga neutra ou bias neutro, que representa uma inclinação ou ponto de vista; bias com carga positiva ou bias positivo, criado para representar as necessidades de informação de um sector cultural ou ideológico; bias com carga negativa ou bias negativo, que se manifesta através de visões ofensivos ou inapropriados. Este trabalho tentará definir essas várias categorias.
Palavras-chave: Auto-bias. Bias relation. Bias positivo. Bias negativo. Bias neutro.

Resumen: El término *bias* es un término-problema en Organización del Conocimiento, principalmente por la falta de una definición clara del mismo. En la literatura del área se pueden encontrar tres aspectos del mismo, *bias* con carga neutra o *bias* neutro, el cual representa una inclinación o punto de vista; *bias* con carga positiva o *bias* positivo, creado para representar las necesidades de información de un sector cultural o ideológico; y *bias* con carga negativa o *bias* negativo, manifestado a través de visiones ofensivas o inadecuadas. El presente trabajo intentará delimitar estos diferentes alcances.
Palavras-clave: Auto-bias. Bias relation. Bias positivo. Bias negativo. Bias neutro.

Abstract: The term bias is a problem term in Knowledge Organization, mainly by the lack of a clear definition. In the literature of the area it is able to find three aspects, bias with neutral charged or neutral bias, which represents an inclination or point of view; bias with positively charged or positive bias, created to represent the information needs of a cultural or ideological sector; bias with negatively charged or negative bias, manifested through visions offensive or inappropriate. This paper will attempt to define these different ranges.
Keywords: Auto-bias. Bias relation. Positive bias. Negative bias. Neutral bias.

Trabalho ganhador do prêmio ISKO Brasil 2015, a ser publicado, na íntegra, na *revista Brazilian Journal of Information Science*.